Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Human Escape Learning in Relation to Reinforcement Variables and Intertrial Conditions: J. H. STRAUGHAN 1

Effects of Stress on Complex Learning and Performance: A. CASTANEDA 9

Effects of Stress and Anxiety on Continuous High-Speed Color Naming: W. Z. DAVIDSON, T. G. ANDREWS, AND S. ROSS 13

Anticipation of Reward as a Function of Partial Reinforcement: H. BRAND, P. J. WOODS, AND J. M. SAKODA 18

Effect of Different Percentages of Money Reward on Extinction of a Lever-Pulling Response: D. J. LEWIS AND C. P. DUNCAN 23

An Analysis of the Effect of Nonreinforced Trials in Terms of Statistical Learning Theory: R. C. ATKINSON 28

A Methodological Study of the Preparation of Connected Verbal Stimuli for Quantitative Memory Experiments: E. E. LEVITT 33

Effect of Overlearning of a Verbal Response on Transfer of Training: G. MANDLER AND S. H. HEINEMANN 39

Simultaneous and Successive Discrimination in a Single-Unit Hollow-Square Maze: A. D. CALVIN AND C. M. WILLIAMS 47

Stimulus Units and Range of Experienced Stimuli as Determinants of Generalization-Discrimination Gradients: J. L. GEWIRTZ, L. V. JONES, AND K. E. WAERNERYD 51

Habit Strength as a Function of Drive in a Brightness Discrimination Problem: E. EISMAN, A. ASIMOW, AND I. MALTZMAN 58

American Psychological Association

Vol. 52 No. 1 July 1956

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1956 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1956 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 52, NO. 1 JULY, 1956

HUMAN ESCAPE LEARNING IN RELATION TO REINFORCEMENT VARIABLES AND INTERTRIAL CONDITIONS [1]

JAMES H. STRAUGHAN [2]

Indiana University

The present experiment was designed to study human escape learning in a simple instrumental situation. In an escape learning procedure, S was required on each trial to choose one of two responses, with correction allowed, in order to turn off an auditory stimulus. A feature of such a design is that it meets the requirements for analysis in terms of the statistical learning theory that has been applied to verbal conditioning experiments (**3, 5**); therefore we can test the extension of the theory to another situation. Finally, we may study the effects of stimulus intensity, a variable which may be important in escape learning although it is not taken into account in the statistical theory, and also certain questions concerning intertrial conditions raised by the verbal conditioning experiments.

The statistical theory, as applied to verbal conditioning, is intended for situations in which (*a*) the organism undergoes repeated trials with the same situation, as experimentally defined, being presented on every trial; (*b*) the responses available to the organism are categorized by experimental criteria as members of mutually exclusive, exhaustive response classes; and (*c*) reinforcement is such that each trial has a fixed probability of terminating with a member of a given response class. The data should be expected to correspond to the model in two major features. First, the probability of responding should change in accordance with a negatively accelerated curve running from the initial probability to the final asymptote. Second, the asymptote which is approached should be equal to the probability that the response will be reinforced when it occurs. These characteristics have appeared in animal T-maze studies (**4**) and verbal conditioning experiments (**3, 5**). A third feature has been found in the verbal conditioning experiments to which the

[1] This study was taken from a dissertation submitted to the Psychology Department of Indiana University in partial fulfillment of the requirements for the Ph.D. degree in psychology. The writer is greatly indebted to Professor W. K. Estes under whose direction the investigation was conducted.

[2] Now at Montana State University.

theory has been applied: the rate of approach to asymptote appears to depend upon the difference between initial response probability and asymptotic probability, with the faster rates being associated with the larger differences. On the hypothesis, proposed elsewhere (**3**), that this third feature depends upon variables associated with massed trials, we should expect that the degree to which it is manifested in the present experiment will be a function of the isolation of trials from one another. Although it may not be possible to achieve an experimental situation in which trials are strictly independent, some control may be exerted over the degree to which trials are isolated from one another by requiring *S*s to perform a distracting task between trials.

Method

Design.—The experiment used escape learning with a tone as the reinforcing stimulus. On each trial, only one of the two possible responses would terminate the tone. The dependent variable was the relative frequency with which one of the response classes, A_1, occurred first on each trial. Trials on which the A_1 response was correct are termed E_1 trials; others are termed E_2 trials. All *S*s received 160 trials, divided into two series, A and B, of 80 trials each. Three independent variables with two values of each variable made an eight-group factorial design appropriate, as summarized in Table 1. The independent variables under investigation were (*a*) the probability that an A_1 response would be followed by termination of the tone on any trial (70% or 100% in Series A, 30% in Series B), (*b*) intertrial conditions (distraction present or absent), and (*c*) the intensity of the tone (80 or 125 db).

TABLE 1

Experimental Design

Group	Probability of Reinforcement (π) Series A	Stimulus Intensity	Distracting Task
I	.70	125 db	Absent
II	.70	125 db	Present
III	.70	80 db	Absent
IV	.70	80 db	Present
V	1.00	125 db	Absent
VI	1.00	125 db	Present
VII	1.00	80 db	Absent
VIII	1.00	80 db	Present

Note.—Probability of reinforcement during Series B was .30 for all groups.

In order to control for possible effects due to differences in sequences of E_1 and E_2 trials, all *S*s within a group were given different reinforcement sequences, although the same 16 random sequences were used for all groups. In randomizing sequences, the probability π of an E_1 trial was fixed, although the proportion of E_1 trials in a series was not. The particular response to be designated A_1 was chosen at random for the first *S* in each group and alternated for subsequent *S*s.

Apparatus.—A sound-insulated room was used with a small fan to provide a masking noise. Illumination was from a 40-w. centrally located bulb. To *S* the apparatus appeared as a gray panel, 35 in. wide by 24 in. high. The auditory stimulus, an impure tone of 500 cycles, was produced by an Army surplus signal generator and presented to *S* by an Army surplus headset. Neither of these pieces of equipment had a manufacturer's identification. The impurities of the tone were not analyzed, although they seemed to be largely harmonics. The soft tone and the loud tone were respectively 80 db and 125 db (re .00025 dynes/cm^2). On the front of the apparatus and 16½ in. above the table top was a black knob 2½ in. in diameter with which *S* could turn off the tone when it came on. This knob was kept centered by a small spring and friction device and required a 70° turn either to the left or to the right in order to turn off the tone. The force required for turning the knob was 20 lb. applied tangentially to the circumference of the knob.

Directly below the knob on the apparatus panel was a horizontal slot through which *S* could see a small black spot when a 7-w. pilot light was on. Also in the slot but in front of the black spot was a small vertical pointer which *S* could move from side to side by means of two handles. When the black spot moved, *S* could keep the pointer in front of it by moving the two handles back and forth. This tracking task is the distracting task mentioned elsewhere in this paper. The slot was 10½ in. above the table top, $\frac{1}{16}$ in. high, and 4½ in. wide. The moving black spot was actually a $\frac{1}{16}$-in. wide, irregular, wavelike black line drawn freehand on a roll of white paper. This paper was moved behind the slot by an Esterline-Angus chart drive so that the line appeared to *S* as a moving black spot. Between the paper (about $\frac{1}{16}$ in. behind the slot)

and the slot was the vertical pointer, made of Brown and Sharpe No. 30 piano wire, which was fastened directly to a bar joining the two controlling handles. These handles, of ½ in. clear plastic rod, were 19¼ in. apart, 8¼ in. above the table top, and extended 1 in. through slots in the front of the apparatus. They were so placed that *S* could rest his arms on the table while controlling them. Behind the panel and out of sight of *S* during the experiment were devices with which *E* controlled the apparatus.

Subjects.—The 128 *S*s were obtained from various spring and summer session classes on the Indiana University campus. Volunteers for the most part, they were right-handed males and females of normal college age and over. They were assigned randomly to the experimental groups.

Procedure.—Each *S* given the distracting task was told that the study concerned the effects of a periodic interruption on tracking. He was told to keep both hands on the handles con-

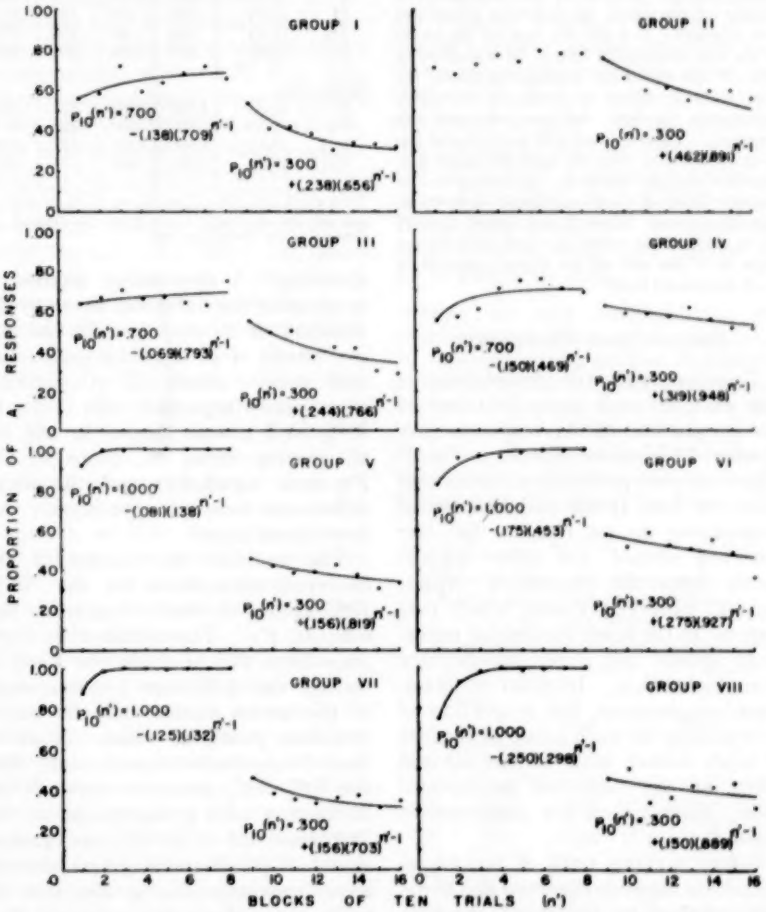

FIG. 1. Observed and theoretical proportions of responses per block of 10 trials (n'). The theoretical curve is in terms of $p_{10}(n')$, the average probability of an A_1 response in the n'th block of 10 trials. A theoretical curve for Group II, Series A, was not obtainable by the methods used for the other groups and series.

trolling the pointer except when he heard the sound in the earphones. When he heard the sound, he was to turn it off with his right hand. The instructions for the groups which did not have the distracting task were similar. They were to keep their hands at rest on the handles during the experiment except when they were turning off the sound.

The trials began approximately 30 sec. after *E* and *S* were prepared. Thereafter, the sound came on for each *S* every 15 sec. for 160 trials without pause. To keep *S*s from long delays in turning off the sound, the following procedures were followed. If *S* did not turn off the sound on the first trial within 15 sec., he was asked to turn off the sound by turning the knob. In some cases *S*s failed to make the corrective response on the first trial that correction was necessary. These *S*s were told at the end of the 15-sec. interval to turn the knob the other way if the first way did not work. In no case was an *S* given either of these additional instructions more than once. When *S* took longer than 15 sec. to turn off the sound, the next trial did not begin until the end of the 15-sec. interval in which he turned it off.

Results and Discussion

Terminal response probabilities.— The data for each group in terms of the proportion of A_1 responses per block of 10 trials are plotted in Fig. 1. There are two predictions concerning observed final levels and theoretical asymptotes to be tested: (*a*) the observed should not differ significantly from the theoretical values, and (*b*) observed values which correspond to the same theoretical asymptote should not differ significantly from each other. In order to make these comparisons, the proportion of A_1 responses for each group in the last 20 trials of each series (A and B) was taken as the observed asymptotic value; these values are presented in Table 2.

Before making tests of the correspondence between observed and theoretical values, a statistical check is called for in order to determine whether or not the groups have arrived at stable asymptotes of responding. A convenient method is to examine the difference between the numbers of A_1 responses in the last two blocks of 10 trials for each series and group. Analyses of variance were made separately for (*a*) the 70%–30% groups, Series A, and (*b*) all groups, Series B. None of the *F*'s were significant and the mean differences were not significantly different from zero.

The analysis of variance of the observed asymptotes for the 70%–30% groups in Series A gave no significant *F*'s. The within-cells variance from this analysis was used in testing the difference between each of the group values and the corresponding predicted value. None of these *t*'s approached significance. For the 70%–30% groups in Series B the distracting task is significant at the .001 level ($F = 16.92$); the groups that had the distracting task showed more responses during the last 20 trials than the other groups. The *t*'s, computed as above, show that the proportions of responses in the last 20 trials of Groups II ($t = 4.98$,

TABLE 2

Mean and Variance of Proportions of A_1 Responses in the Last 20 Trials of Each Series

Group		Series A		Series B	
		Mean	Variance	Mean	Variance
70%–30%					
I	(LA)*	.688	.038	.332	.040
II	(LP)	.788	.032	.581	.075
III	(SA)	.684	.040	.294	.021
IV	(SP)	.697	.054	.509	.067
100%–30%					
V	(LA)	1.000	.000	.322	.007
VI	(LP)	.997	.000	.425	.065
VII	(SA)	1.000	.000	.328	.020
VIII	(SP)	1.000	.000	.372	.046

* *L* and *S* stand, respectively, for the loud (125 db) and soft (80 db) tone; *A* and *P* for the absence and presence of the distracting task.

$P = .001$) and IV ($t = 3.70$, $P = .01$), the groups with the distracting task, are significantly above the predicted asymptotic values. The analysis of variance of the observed asymptotes for the 100%–30% groups in Series B showed none of the F's to be significant, although computation of the t's as above showed that the proportion of responses in the last 20 trials of Group VI is significantly above the predicted asymptotic value ($t = 2.70$, $P = .02$). No tests were made of the asymptotes of the 100%–30% groups, Series A, since it is obvious that they conform closely to the predicted values.

We may conclude that the observed asymptotes behave as expected in Series A but that there are some deviations in Series B. These deviations occur predominantly in those groups receiving the distracting task and are all above the theoretical asymptote. The interpretation may be either that the frequency curves are approaching a higher asymptote or that they are approaching the predicted asymptote but at a slower rate. In favor of the latter interpretation, we note that those groups that show a significant deviation from the predicted asymptote in Series B are still decreasing in the last two blocks of trials.

Learning rates.—Differences in response totals may be taken as evidence for differences in learning rates when both initial response levels and asymptotes can be assumed constant in the given comparison. Since much of the learning seems to have occurred early in the series, the comparison of learning rates was made by using mean numbers of responses in the first 20 trials rather than means taken over the total series. The mean number and the variance of A_1 responses made by each group in the first 20 trials of each series are shown in Table 3. Analyses of variance for the 70% and 100% groups were done separately for Series A and again for Series B. For the 100% groups in Series A the data were transformed by the formula, $y = e^{1.2x}$, because Bartlett's test indicated that the variances were not homogeneous. The significant results of these analyses were as follows: (*a*) the interaction between sound intensity and distracting task is significant for the 70%–30% groups in Series A ($F = 6.09$, $P = .05$); (*b*) the effect of the distracting task is significant for the 70%–30% groups in Series B ($F = 9.01$, $P = .01$), and for the 100%–30% groups in Series *A* ($F = 11.1$, $P = .001$) and Series B ($F = 4.01$, $P = .05$).

Thus, we may conclude that the distracting task increased the number of A_1 responses during the first 20 trials for all groups in Series B, and decreased the number of A_1 responses during the first 20 trials for the 100%–30% groups in Series A. Both of these effects were in the direction of slower learning for the groups with the distracting task. The

TABLE 3

MEAN NUMBER AND VARIANCE OF A_1 RESPONSES IN THE FIRST 20 TRIALS OF EACH SERIES

Group		Series A		Series B	
		Mean	Variance	Mean	Variance
70%–30%					
I	(LA)*	11.50	11.07	9.50	18.13
II	(LP)	14.00	10.93	14.24	15.97
III	(SA)	12.93	7.60	10.32	18.86
IV	(SP)	11.19	17.97	12.00	21.07
100%–30%					
V	(LA)	19.13	.58	8.75	10.73
VI	(LP)	17.81	8.12	10.94	14.67
VII	(SA)	18.63	1.25	8.31	10.67
VIII	(SP)	16.81	19.85	9.50	9.33

* See Table 2 for meaning of letter symbols.

trend for the 70%–30% groups, Series A, was not consistent. The interpretation of the significant interaction between effects of the distracting task and sound intensity in the 70%–30% groups, Series A, is made difficult by the possible differences in starting points, i.e., differences in $\bar{p}(0)$. There is no evidence that the independent variable of sound intensity had any effect upon the learning rate aside from this interaction effect.

Parameter estimates.—An analysis of the results further in terms of statistical learning theory requires reference to aspects of the model that are applicable to the present experiment. From assumptions discussed elsewhere (**1, 2, 3**) the following equation for response probability as a function of trials can be derived:

$$\bar{p}(n) = \pi - [\pi - \bar{p}(0)](1 - \theta)^n, \quad (1)$$

where $\bar{p}(n)$ is the mean probability of an A_1 response on trial $n + 1$ and θ is a parameter describing learning rate, or, in terms of statistical learning theory, θ is the proportion of stimulus elements sampled on each trial.

In order to account for the possibility that successive samples of stimulus elements are not independent, we may introduce two parameters, u and b to replace θ, where u is the average probability that an element is in the sample on a trial and b is the probability with which an element remains in the sample from the preceding trial. We should expect that insertion of a distracting task between trials should decrease the size of u, the size of b, and the difference between b and u, with b being larger than u. It can be shown (**6**) that, with the additional parameter, Equation 1 takes the form

$$\bar{p}(n) = \pi - [\pi - \bar{p}(0)] \times (1 - b)(1 - u)^{n-1}. \quad (2)$$

Equation 2 can be put into the more convenient form

$$p_{10}(n') = \pi - [\pi - p_{10}(1)] \times (1 - u)^{10(n'-1)}, \quad (3)$$

where $p_{10}(n')$ is the average probability of an A_1 response in the n'th block of 10 trials.

TABLE 4

ESTIMATES OF u AND b FOR EACH SERIES

Group		u		b	
		Series A	Series B	Series A	Series B
70%–30%					
I	(LA)*	.034	.041	.23	.20
II	(LP)	—†	.012	.18	.09
III	(SA)	.023	.026	.27	.33
IV	(SP)	.073	.005	.11	.11
100%–30%					
V	(LA)	.180	.020	—	.36
VI	(LP)	.076	.008	—	.28
VII	(SA)	.183	.035	—	.35
VIII	(SP)	.114	.012	—	.23

* See Table 2 for meaning of letter symbols.
† Value of u for Group II, Series A, not obtainable from data.

The values of u obtained by applying Equation 3 are given in Table 4. For algebraic reasons no solution is obtainable for Group II, Series A; therefore no theoretical curve is fitted to this data in Fig. 1. Also, values of u are not obtainable for many individual *S*s; we cannot, therefore, use analyses of variance based on individual u's to test group differences. An alternative method of testing the significance of effects of experimental variables upon u values is to compare groups with respect to learning rates for each of the series separately. This has been done in the preceding section and the results confirm predictions concerning the size of u.

The values of b obtained for each group by finding the difference between the conditional probabilities (**3**) are shown in Table 4. Because E_2 trials did not occur during Series A for the 100%–30% groups, values of b for these groups are not obtainable from the data. Our expectation that b should be smaller for the groups that had the distracting task is confirmed by Table 4. Analyses of variance were computed for the 70%–30% groups, Series A and B, as shown

TABLE 5

ANALYSES OF VARIANCE OF b AND (b-u) VALUES FOR THE 70%–30% GROUPS, SERIES A AND B

Source	*df*	b		(b-u)	
		Mean Square	*F*	Mean Square	*F*
Between groups	3	.2451	3.358*	.2482	3.301*
Sound intensity	1	.0241	.330	.0116	.154
Distracting task	1	.6052	8.290**	.5468	7.271**
S.I. × D.T.	1	.1059	1.451	.1863	2.477
Between *Ss*, within groups	60	.0730		.0752	
Series	1	.0100	.382	.0000	.000
Series × Groups	3	.0305	1.164	.0438	3.532*
Series × S.I.	1	.0578	2.206	.1203	9.702**
Series × D.T.	1	.0336	1.282	.0063	.508
Series × D.T. × S.I.	1	.0000	.000	.0049	.395
Ss × Series, within groups	60	.0262		.0124	

* Significant at .05 level.
** Significant at .01 level.

in Table 5, and for all groups, Series B, as shown in Table 6. The effect of the distracting task is significant at the .01 level for the 70%–30% groups in Series A and B ($F = 8.29$) and for all groups in Series B ($F = 10.00$). The groups that had the distracting task showed smaller values of b. In the analyses of Table 5, the between-*Ss* within-groups variance estimate was used as the error term in computing the between-groups *F*'s, and the *Ss*-by-series within-groups variance estimate was used as the error term for the series and series-by-groups *F*'s. The size of b appears to be independent of sound intensity and series; however, there is a significant difference between the 70%–30% groups and the 100%–30% groups in Series B.

Another indication of the degree of independence of stimulus samples drawn on successive trials should be the size of differences between b and u. If b and u are equal, then successive samples are presumably statistically independent, and the mathematical model we are using reduces to the simpler model where θ is equal to b and u. If, then,

TABLE 6

ANALYSES OF VARIANCE OF b AND (b-u) VALUES FOR ALL GROUPS, SERIES B

Source	*df*	b		(b-u)	
		Mean Square	*F*	Mean Square	*F*
Within groups	120	.0583		.0584	
Between groups	7	.1827	3.134**	.1647	2.837*
Sound intensity	1	.0123	.211	.0134	.229
Distracting task	1	.5827	9.995**	.4122	7.058**
Reinforcing conditions	1	.4995	8.568**	.5432	9.301**
S.I. × D.T.	1	.0517	.887	.0488	.836
S.I. × R.C.	1	.0805	1.381	.0922	1.579
D.T. × R.C.	1	.0397	.681	.0078	.134
S.I. × D.T. × R.C.	1	.0123	.211	.0421	.721

* Significant at .05 level.
** Significant at .01 level.

(b-u) is smaller for those groups having the distracting task, we may accept this as evidence supporting the mathematical theory. In order to examine the effect of the distracting task upon the size of (b-u), individual values of (b-u) were obtained by subtracting the value of u obtained for a group from the value of b obtained for each individual S in the group. Analyses are shown in Tables 5 and 6. The error terms used in the analysis of (b-u) in Table 5 are the same as those used for b. The effect of the distracting task is significant at the .01 level for the 70%–30% groups in Series A and B ($F = 7.27$) and for all groups in Series B ($F = 7.06$); the groups with the distracting task showed smaller values for (b-u) than the groups without the distracting task. Not predicted in the model are the significant interaction between series and sound intensity in the 70%–30% groups, and the significant difference between the 100%–30% groups and the 70%–30% groups in Series B.

Summary

An experiment was conducted to study two-choice escape learning in a simple instrumental situation with human *Ss*. The experimental design grew out of an analysis of the situation in terms of the statistical learning theory that has been applied to verbal conditioning experiments (**3, 5**).

Eight groups of 16 *Ss* each were used in a three-variable factorial design. Two mutually exclusive responses were available to *S* on each trial, turning a knob clockwise and turning it counterclockwise; reinforcement was cessation of a tone following choice of the correct response. All *Ss* received 160 uninterrupted trials spaced at 15-sec. intervals. The independent variables were probability of reinforcement, presence or absence of a distracting task between trials, and intensity of the tone. It was found that terminal response probability tended in most cases to match probability of reinforcement; learning was generally slower with the distracting task; and stimulus intensity had little effect. Predictions made on the basis of statistical learning theory appear to be largely verified.

References

1. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
2. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
3. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
4. Lauer, D. W., & Estes, W. K. Observed and predicted terminal distributions of response probability under two conditions of random reinforcement. *Amer. Psychologist*, 1954, **9**, 413. (Abstract)
5. Neimark, E. D. Effects of type of non-reinforcement and number of alternative responses in two verbal conditioning situations. Unpublished doctor's dissertation, Indiana Univ., 1953.
6. Straughan, J. H. An application of statistical learning theory to an escape learning situation using human subjects. Unpublished doctor's dissertation, Indiana Univ., 1953.

(Received July 22, 1955)

Journal of Experimental Psychology
Vol. 52, No. 1, 1956

EFFECTS OF STRESS ON COMPLEX LEARNING AND PERFORMANCE

ALFRED CASTANEDA[1,2]

Iowa Child Welfare Research Station, State University of Iowa

In a previous study (**2**) performance on a second task was studied as a function of whether the S-R pairs to be relearned were the same or different from those learned in the first task and of instructions emphasizing speed of responding. A significant interaction was found based on poorer performance under the speed condition on those S-R pairs which were different and a tendency for this condition to facilitate performance on those pairs which were the same.

Assuming that the correct response tendency is dominant where the S-R pairs are the same while the incorrect tendency is dominant in the case where they are different, these results were interpreted to be in accord with Spence's (**8**) derivations from Hull (**4**) regarding performance in complex learning situations as a function of variations in the level of drive. On the basis of these derivations, and assuming that the speed instructions served to increase the drive level, it would be expected that performance should be deleteriously affected by an increase in the drive level where the incorrect tendency is dominant. However, facilitation would be expected where the correct response tendency is dominant.

The present study provides further data relevant to these expectations. Dominance of either the correct or incorrect response tendencies is assumed to be varied in the learning of a second task by a modification of a procedure employed by Duncan (**3**) and Porter and Duncan (**7**). That is, the second task requires learning to re-pair half of the S-R pairs which have been learned in the first task while the other half is left unchanged. The method for varying the level of drive differs from the previous study in that a modified task pacing technique is used. It has been assumed by others (**5, 6**) that stress is a result of excessive task pacing. In the present study it is assumed that this method of inducing stress results in variations in the level of drive.

METHOD

Subjects.—A criterion of less than 15 errors on the first task was established, on the basis of preliminary work, as a basis for selecting *Ss* in the present study. Sixty-two fourth- and fifth-grade *Ss* from a total of 78 met this criterion. These *Ss* were randomly assigned in equal numbers to the paced (stress) and nonpaced (nonstress) conditions.

Apparatus and procedure.—The apparatus consisted of a box 13 × 12¼ × 5¼ in., painted flat black. Projecting from its base was a sloping response panel approximately 7¼ × 9¼ × 13 in. Two push buttons were arranged on this panel on a horizontal plane spaced 6 in. apart and 2 in. from *S*'s edge of the panel and toward its center. On the panel and so that it formed the apex of a triangle with the two push buttons serving as the points connecting the base was a green, jeweled reflector illuminated by a 6.3-v. pilot lamp. Centered 4 in. above the response panel was a 1 in. diameter aperture of flashed opal glass. Housed behind the aperture were four 6.3-v. lamps three of which had been colored red, blue, or amber. The construction of the apparatus was such that by means of a

[1] Appreciation is expressed to Robert Everetts, City Superintendent of Schools, Helen Levis, Principal, Lincoln Elementary School, and Elsie Glassburner, Principal, Stewart Elementary School, Washington, Iowa for their assistance and cooperation in providing experimental facilities and arranging for subjects.

[2] These data were collected by Lewis Lippsit.

TABLE 1

ERRORS TO CHANGED AND UNCHANGED PAIRS ON SECOND TASK

Group	S-R Pair			
	Changed		Unchanged	
	Mean	*SD*	Mean	*SD*
Stress	9.58	6.27	4.35	3.60
Nonstress	7.32	4.15	5.77	3.17

rotary selector switch any single light could be activated and by means of a toggle switch either one of the two buttons could be set to turn it off. If the correct button was selected, the light went off and the green light on the response panel was activated. Incorrect responses affected neither the stimulus light nor the green light on the response panel. During the second task the amount of time the stimulus light was on was controlled by means of an electronic interval timer. The arrangement, four lights and two buttons, permitted construction of a task composed of four S-R pairs. Each button was learned to two of the lights separately. The selection of the two pairs to be re-paired in the second task was determined randomly with the exception that it result in each of the two buttons being associated with two of the lights.

With the restriction that each light appear once in every four presentations, the order of presentation was randomly determined for both tasks. All *Ss* were given 8 presentations of each light in the first task for a total of 32, and 10 of each light in the second task for a total of 40 presentations. In the first task a light remained activated until *S* made the correct response. However, in the second task only one response per light was permitted. No attempt was made to control the interval between stimulus presentations in both tasks.

The instructions for the first task were the same for all *Ss*. They were designed to indicate to *S* that the task required learning which of the two buttons turned off a given colored light. Upon completion of the first task a 2-min. rest period was introduced. Other than to require *S* to remain seated no other attempt was made to control his activity during this period. At the end of the rest period all *Ss* were told that they would not be permitted to correct for errors. No information was given regarding the fact that half of the S-R pairs would require re-pairing. In addition to these instructions *Ss* assigned to the stress condition were given a demonstration in which two lights were activated for 1 sec. each and informed that failure to respond within this interval would be scored as an error. However, during the task the timer was set to give a 2-sec. exposure of the light and if a response was not made during this interval, the same light was re-presented until a response could be recorded for that trial (failures to respond do not constitute a part of the data in this study). In no case, however, did a given light have to be re-presented more than once. Only four *Ss* in the stress condition failed to respond within the 2-sec. interval on the first presentation of a light. For these *Ss* a light selected for re-pairing had to be re-presented a total of three times and a light which was not to be re-paired for a total of three times also.

RESULTS AND DISCUSSION

Error scores of the two groups for the 32 trials in the first task were subjected to a *t* test to test the hypothesis that they were random samples from the same population. The resulting *t* of .23 with 60 *df* permitted retention of the null hypothesis.

Table 1 presents the number of errors made by each group on each type of S-R pair separately in the second task. On the basis of these

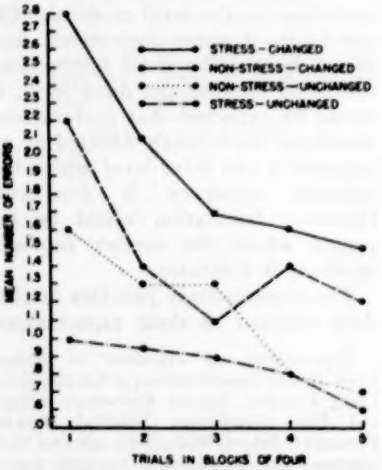

FIG. 1. Performance curves on each type of S-R pair separately for stress and nonstress.

data it is apparent that the stress group performed better than the nonstress group on the unchanged pairs but more poorly on the changed pairs.

Figure 1 presents these data in blocks of four presentations for each type of pair separately. With the exception of the fourth block, it can be seen that the performance of the stress group is consistently superior to that of the nonstress group on the unchanged pairs and inferior, without exception, on the changed pairs.

Table 2 presents a summary of an analysis of variance based on the data presented in Fig. 1. This analysis indicates that the main effects of stress are not significant. The effects of re-pairing as well as those due to blocks are highly significant. The significant interaction between Blocks and Pair, while not of particular interest in the present study, is apparently due to the overlapping between the performance curves of the changed and unchanged pairs of the nonstress group. It is the significant interaction between Stress and Pair, however, that is of central importance. Viewed in conjunction with the data presented in Fig. 1, this interaction may be interpreted as indicating that the effects of stress are contingent on the type of S-R pair involved. That is to say, stress results in fewer errors, in comparison to nonstress, on the unchanged pairs and in more errors on the changed pairs. Since the triple interaction did not indicate that the magnitude of such effects varied with blocks, tests of the simple effects of stress are based on the means shown in Table 1. These indicated that the difference in errors on the unchanged pairs was significant at only the .20 level and at the .10 level on the changed pairs.

TABLE 2

ANALYSIS OF VARIANCE OF ERROR SCORES ON SECOND TASK

Source	*df*	*MS*	*F*
Stress	1	.93	—
Error (b)	60	4.60	
Blocks	4	15.81	15.65**
Pair	1	71.13	18.26**
Blocks × Pair	4	3.65	4.74**
Blocks × Stress	4	.20	—
Pair × Stress	1	21.70	5.58*
Blocks × Pair × Stress	4	1.25	1.62
$Error_1$ (w)	240	1.01	
$Error_2$ (w)	60	3.89	
$Error_3$ (w)	240	.77	

* $P < .025$.
** $P < .001$.

One of the properties commonly ascribed to drive states (**1**) is that they serve to combine with and enhance the strength of those response tendencies aroused by the stimulus situation. Moreover, Spence (**8**) has outlined the theoretical considerations which necessitate control, in those situations where more than one response tendency is aroused, of whether the dominant tendency at the time drive is increased is correct or incorrect for the task. Other things being constant, and provided this factor can be specified, either improvement or impairment of performance may be predicted as a function of stress on the assumption that it serves to increase the level of drive. In the present situation, then, it would be expected that stress should facilitate performance on the unchanged pairs and impair it on the changed pairs. The results did tend to confirm these expectations.

A variety of methods have been used for varying stress in complex learning situations and a frequent finding is for performance to be impaired under stress (**5**). It is quite reasonable to presume that the different methods employed contribute their own unique set of stimuli to the total stimulus complex. Associated with such stress-produced stimuli are previously learned or unlearned responses which may be antagonistic and/or incompatible with

those required for correct performance. Hence, quite aside from its assumed intensifying property, should such incompatible tendencies become dominant in the situation, it would seem reasonable to expect general impairment of performance under stress. Since the present results indicated that stress tended to facilitate performance on the unchanged pairs, it may be assumed that this factor was controlled in the present study.

Summary

The present study was concerned with performance in a second task as a function of task pacing and whether the S-R pairs to be relearned were identical to those learned in the first task or were required to be re-paired.

All 62 *Ss* were first given 32 trials on the first task. Half of the *Ss* relearned the second task under conditions of excessive task pacing. The second task required re-pairing of half of the S-R pairs learned in the first task while the other half remained unchanged. A significant interaction between these two variables was found. While the simple effects were not statistically significant they did tend to indicate that the stress condition resulted in fewer errors on the unchanged pairs in comparison to nonstress and in comparatively more errors on the changed pairs.

References

1. Brown, J. S. Problems presented by the concept of acquired drives. In *Current theory and research in motivation: A symposium.* Lincoln, Neb.: Univ. of Nebraska Press, 1953.
2. Castaneda, A., & Palermo, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, **50,** 175–179.
3. Duncan, C. P. Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *J. exp. Psychol.*, 1953, **45,** 1–11.
4. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
5. Lazarus, R. S., Deese, J., & Osler, S. F The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49,** 293–317.
6. Lindsley, D. B. *A study of performance under speed stress.* (OSRD, 1945, Publ. B. No. 1838.) Washington, D. C.: U. S. Dep. of Commerce, 1946.
7. Porter, L. W., & Duncan, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **45,** 91–96.
8. Spence, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *The Kentucky Symposium: Learning theory, personality theory, and clinical research.* New York: Wiley, 1954.

(Received July 8, 1955)

Journal of Experimental Psychology
Vol. 52, No. 1, 1956

EFFECTS OF STRESS AND ANXIETY ON CONTINUOUS HIGH-SPEED COLOR NAMING[1]

WILLIAM Z. DAVIDSON,[2] T. G. ANDREWS, AND SHERMAN ROSS

University of Maryland

The effects of stress on performance is a complex and important problem for psychological analysis. Many studies indicate that experimentally induced stressful conditions produce decrements in performance (**1, 3, 4, 11**). The results of other studies indicate that no decrement is produced under stress (**7, 17**). Still others have found both decrements and increments in performance as a result of stress (**7, 13, 17**). Two main sources have been suggested as important factors affecting the relationship: (*a*) the method of experimentally inducing stress and (*b*) the task to be performed. Closely associated with the first source is the problem of motivation.

This study has three major objectives: (*a*) to compare two different methods of inducing stress, (*b*) to explore the relationship between anxiety and performance under stressful conditions, and (*c*) to study stress and anxiety effects on the performance of a task of relative generality.

Two methods of inducing stress were selected with regard to the classification reported by Lazarus, Deese, and Osler (**12**). The first method, *failure stress*, was induced by reporting false failure scores to the Ss. The second method, *task-induced stress*, was produced by the rapid presentation of stimuli for discrimination. The need for comparative studies of the different methods for experimentally inducing stress has been pointed out by Lazarus, Deese, and Osler (**12**) and by Holtzman and Bitterman (**6**).

In recent years an increasing interest has been developing in the attempt to account for some of the differential results of performance under stress by seeking relationships between manifest anxiety and performance (**10, 13**). The present experiment also deals with this characteristic.

METHOD

Design.—A basic 3 × 3 × 2 factorial design was used with three levels of failure stress, three levels of task-induced stress, and two levels of anxiety, with three *Ss* per cell. This design was used in a manner developed by Johnson and Tsao (**8**) to assess the factor of individual differences.

Task.—The performance task selected for the experiment was continuous, high-speed color naming, which was considered as representative of a wide variety of performances involving continuous visual discrimination. A continuous 70-unit series of five different colors in random order on a Kodachrome film strip was projected in *S*'s view. The colors were red, green, pink, yellow, and blue. The *S*'s task was to name each color as it appeared.

Variables.—Task-induced stress was generated by varying the speed of presentation of the colors. The three speeds of presentation, automatically controlled, were 75, 85, and 100 stimuli per minute. At these speeds, each stimulus was exposed for .8, .7, and .6 sec., respectively.

Failure stress was induced by verbal instructions and electric shock as an indicator of failure. Before the testing *S* was informed that per-

[1] This experiment is one of a series of studies on behavior decrement performed under Contract No. DA-49-007-MD-222 between the University of Maryland and the Medical Research and Development Board, Office of the Surgeon General, Department of the Army.

[2] This report has been prepared by T. G. A. and S. R. alone due to the untimely death of Mr. Davidson in March, 1954.

formance on the task was related to mechanical aptitude, clerical aptitude, and intelligence test scores. He was also told that the apparatus was set to deliver automatically an electric shock for poor performance at the end of each test period. Each *S* was then given the same moderate shock as a demonstration of what to expect for failure. What constituted "failure," punishable by shock, was different for each of the three groups. The "slight failure" group was told that only if their performance fell *slightly below* the performance level of the average high school senior would they receive the electric shock. The "moderate failure" group was told that only if their performance fell *considerably below* the level mentioned would they receive the shock. The "high failure" group was told that only if their performance fell *extremely far below* the performance level of the average high school senior would they receive the shock.

Each *S* was told that after 30 sec. on the task he would receive an auditory signal which would inform him of the efficiency of his performance up to that point. He was told that if he heard a bell he was doing well enough to avoid the shock. If he heard a buzzer, however, he was in danger of receiving the shock, unless his performance improved throughout the remainder of the trial.

After 30 sec. on the task, the buzzer was sounded for all *S*s, and at the end of the performance period all *S*s received the electric shock. After the 3-min. trial a rest period of 30 sec. was allowed in order to permit the distractive effects of the shock to dissipate. During this period, *E* made derogatory comments about *S*'s performance. The comments were more derisive the higher the failure condition to which the *S* had been assigned. The task was resumed, repeating the performance and shock periods twice for a total of three performance periods and three shocks.

The shock was utilized in the failure stress conditions in an attempt to keep motivation roughly equivalent for all *S*s. A momentary impulse shock of 225 v. was found to be fairly uncomfortable, but not too disrupting. This shock of moderate intensity was used since it was felt that an intense shock would destroy any effect produced by the failure instructions. The *S*s might be concerned only with avoiding the shock and completely ignore the failure instructions. However, a moderately annoying shock would serve to motivate *S* to perform at a high level in order to avoid it, but would not be so disruptive as to destroy the effects of the failure instructions.

The *S*s scoring in the upper and in the lower range on a modified Taylor Scale of Manifest Anxiety constituted the high and low anxiety groups in the experiment. The modified form of the Taylor scale (**16**) had been administered under the innocuous title of a Biographical Inventory to 234 students in an introductory psychology course at the University of Maryland. The distribution of scores was similar to that found by Taylor on a similar population. The mean was 14.50, and the distribution was positively skewed with a range of scores from 2 to 28.

Subjects.—Twenty-seven male students making scores of 8 and below, and 27 making scores of 21 and above were selected for use as the high- and low-anxiety groups. These scores were approximately one *SD* from the mean of the obtained distribution. Within each anxiety group *S*s were assigned at random to the various cells of the factorial design. Each *S* was paid $1.35 for his services in the experiment.

Procedure.—Each *S* was tested with the American Optical Company's Pseudo-Isochromatic Color Perception Test to insure that he was not color deficient. He was seated 2 ft. from a large flat black shield, in the center of which was a $1 \times 1\frac{1}{2}$-in. ground-glass screen placed at eye level and onto which the colors were projected. In front of the shield was an Audograph conference microphone into which *S* called out the color names. Leading from the lower left side of the shield was a wire with an electrode strap. The strap was secured to *S*'s left wrist by means of two brass electrodes.

The *S* was given preliminary instructions about the importance of the task and the discriminations and responses that would be required of him. A practice trial was then given, consisting of one run through the 70-unit color series. The colors were projected from behind the shield through a 1-in. circle onto the ground-glass screen. In the practice trial the colors were changed manually after *S* called out the name of each color as it was exposed at a self-paced rate. When the practice trial was completed, each *S* was given further instructions to establish the appropriate failure condition. The test trials were then begun. In the test trials, the colors changed automatically at the predetermined rate. As *S* called out the colors, a recording of his responses was made.

Each color frame on the film had a code number which was projected onto the ground-glass screen. This code number was projected outside the aperture through which *S* was viewing the stimuli. Only *E* could see the code numbers, and he whispered every fifth code number into a separate microphone. In this manner the code numbers were superimposed on the record of *S*'s responses for later transcription and scoring.

Immediately after the test trials were completed, *S* was given a Stress Experience Inventory to complete. This inventory consisted of four 7-point rating scales and 25 multiple-choice

items relating to the amount of stress *S* recalled having experienced during the experimental session. This inventory was introduced to see if there was a relationship between subjective feelings of stress and both performance and the conditions of stress in the experiment. No further verbal instructions were given until after the inventory was completed. The *E* then described the true nature of the experiment to *S* and indicated that he had actually done well. This was done in order to remove any possible effects of the imposed failure conditions.

Results and Discussion

The transcribed performance records for each *S* were scored in terms of errors of omission (blocking), errors of commission (incorrect response), and total number of correct responses. Preliminary analyses of these scores indicated that errors of commission were relatively infrequent and were less sensitive to variations in the experimental conditions. Total correct responses merely represent the complement of the sum of the two types of errors. Errors of omission—no response to a given stimulus—were used in the major analyses of results.

TABLE 1

Analysis of Variance: Transformed Error Scores After Warning of Potential Failure

Source	Mean	*df*	*MS*	*F*
Speed (S)		2	591.26	38.07†
Slow	12.65			
Medium	18.79			
Fast	24.11			
Failure (F)		2	55.20	3.55*
Slight	17.22			
Moderate	20.50			
High	17.82			
Anxiety (A)		1	34.18	2.20
Low	17.72			
High	19.31			
Ss		2	610.60	39.32†
S × F		4	66.83	4.30**
S × F × A		4	77.82	5.01**
Error‡		38	15.53	
Total		53		

* $P < .05$.
** $P < .01$.
† $P < .001$.
‡ Error term is pooled insignificant interactions.

TABLE 2

Analysis of Variance: Transformed Error Scores After Shock Indication of Failure

Source	Mean	*df*	*MS*	*F*
Speed (S)		2	400.61	16.20**
Slow	5.54			
Medium	9.34			
Fast	14.92			
Failure (F)		2	39.76	1.61
Slight	8.37			
Moderate	11.30			
High	10.14			
Anxiety (A)		1	40.98	1.66
Low	9.06			
High	10.81			
Ss		2	638.00	25.80**
S × F		4	73.98	2.99*
A × *Ss*		2	100.18	4.05*
Error†		40	24.73	
Total		53		

* $P < .05$.
** $P < .001$.
† Error is pooled insignificant interactions.

Two separate combinations of the errors of omission were made: (*a*) for the three "postbuzzer" periods, in which *Ss* were operating under threat of failure; and (*b*) for the "postshock" periods, in which *Ss* were operating under presumed failure. The absolute scores were converted to percentages, since under different speed conditions *Ss* responded to differing numbers of stimuli. The inverse sine or angular transformation (**5**, p. 203) was applied to the percentage data in order to permit application of analysis of variance techniques to the results.

The results of the analyses of the transformed data on errors of omission are presented for the two aspects of performance: after signal for possible failure (Table 1) and after shock indication of failure (Table 2). These tables also present the appropriate means.

There are certain features of these analyses which deserve special comment. Absolute replication is dispensed with as a result of classifying the three individuals in each cell with respect to their rank orders of errors of omission. This arrangement, devised by John-

son and Tsao (**8, 9,** p. 311), places individual order as a main effect. Since there is then no independent comparison attributable to experimental error, the highest order interaction is first used to test the main effects and each of the other interactions. The nonsignificant interactions are pooled as the general error term (**2, 8, 9, 15**). This general technique permits evaluation of the relative significance of individual differences as a separate source of variation in the experimental results. It has seemed to us that this feature is of some importance in considering direct applications of any experimentally obtained results. The data from the Stress Experience Inventory are presented in Table 3. In this analysis the more conventional form of analysis of variance was used.

Another and more limiting feature of the results in Tables 1 and 2 is the fact that the analyses were necessarily performed on transformed scores. There is common effect between the original and the transformed scores, but they are not of exactly the same distribution. This feature necessarily places restrictions on the generalizations that may be made and on the interpretation of interaction effects (**14,** p. 149). This situation is typical of experimental data of the form used in studies of this type.

TABLE 3

ANALYSIS OF VARIANCE: STRESS EXPERIENCE INVENTORY

Source	Mean	*df*	*MS*	*F*
Speed (S)		2	7.91	3.92*
Slow	5.34			
Medium	5.50			
Fast	6.56			
Failure (F)		2	3.85	1.90
Slight	6.17			
Moderate	5.28			
High	5.94			
Anxiety (A)		1	22.68	10.23**
Low	5.15			
High	6.44			
S × F		4	8.94	4.42**
S × A		2	1.68	
F × A		2	9.86	4.88*
S × F × A		4	2.77	1.37
Error (within)		36	2.02	
Total		53		

* $P < .05$.
** $P < .01$.

As seen in the analysis tables, the experimental conditions clearly appear to have produced "stress" as we have defined it. This stress was reflected both in the performance and in subjective reports. On the color-naming task, errors of omission appear to be sensitive indicators of stress. Task-induced stress and failure stress conditions appear to produce effects upon performance, the former exhibiting a monotonic relationship and the latter a nonmonotonic relationship. The results also suggest that failure stress may dissipate in time. Cognitive processes may intercede to reduce the effects of failure stress.

Regarding the role of manifest anxiety, it appears from the results of this study that anxiety classification as such does not produce differential effects on the continuous high-speed color-naming performance. However, the several significant interactions obtained between anxiety and the experimental conditions indicate a possible "priming" function of anxiety for other stress effects. As might be expected, anxiety served to increase *S*'s susceptibility to the effects of stressful conditions. This interpretation is supported by the fact that anxiety conditions of the *S*s produced significant differences on the responses to the Stress Experience Inventory. These results for the inventory serve to substantiate a basic assumption upon which the study was based, to the effect that the experimental conditions imposed upon the *S*s would serve as stressful experiences.

The highly significant results for the individual orders obtained in the experiment clearly indicate that there are very wide individual differences in the performance of the *S*s on this experimental task. The interactions obtained as statistically significant are of value here in generating hypotheses for further investigation, specifically designed effectively to test these hypotheses. The speed-failure interaction obtained in the three analyses is of particular interest. Examination of the tabulated means indicated that

high speed and failure stress, especially the "moderate failure" condition, combined to produce a disproportionately large number of errors of omission, a form of blocking. The combined stress here may well be producing a characteristic disruption which can be tested separately. The interactions with the anxiety factor are more easily characterized as indicating a heightened sensitivity to the experimental stress on the part of the more "anxious" *Ss*, as indicated by analysis of the tabulated means.

Summary

This experiment was designed to determine the differential effects of anxiety, failure stress, and task-induced stress on performance of a continuous high-speed color-naming task. A total of 54 male undergraduate students were selected as *Ss* on the basis of high and low scores on a modified Taylor Anxiety Scale. They were randomly assigned within each anxiety group to the various cells of a basic $3 \times 3 \times 2$ factorial design. This design involved three levels of failure stress, three levels of task-induced stress, and two levels of anxiety classification. The design permitted the assessment of individual differences as a separate source of variation among the results. Errors of omission were used as measures of performance.

Performance scores were studied for two special portions of the experimental data: (*a*) after warning of potential failure and (*b*) after shock indicating failure. A Stress Experience Inventory was given to each *S* after his testing session.

The results show that errors of omission are a sensitive indicator of stress effects. Failure stress and task-induced stress produce significant decrements in continuous high-speed color-naming performance and significant effects in subjective reports of stressful experience. Manifest anxiety does not produce differential effects on performance. The several significant interactions between anxiety and the various experimental conditions indicate a possible "priming" function of anxiety. Highly significant individual differences were found for the task used in the experiment.

References

1. Bills, A. G., & Shapin, M. J. Mental fatigue under automatically controlled rates of work. *J. gen. Psychol.*, 1936, **15**, 335–347.
2. Binder, A. The choice of error term in analysis of variance designs. *Psychometrika*, 1955, **20**, 29–50.
3. Conrad, R. The effects of very fast speeds on multiple dial watching. *Med. Res. Coun., Appl. Psychol. Unit*, 1949, No. 115.
4. Conrad, R. Speed and load stress in a sensori-motor skill. *Brit. J. indus. Med.*, 1951, **8**, 1–7.
5. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
6. Holtzman, W. H., & Bitterman, M. E. Psychiatric screening of flying personnel. VI. Anxiety and reactions to stress. *USAF Sch. Aviat. Med., Proj. Rep.*, 1952, Proj. No. 21-37-002 (Rep. No. 6).
7. Hurlock, E. B. The value of praise and reproof as incentives for children. *Arch. Psychol.*, N. Y., 1924, **11**, No. 71.
8. Johnson, P. O., & Tsao, F. Factorial design and covariance in the study of individual educational development, *Psychometrika*, 1945, **10**, 133–162.
9. Johnson, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.
10. Krugman, H. E. Flicker-fusion frequency as a function of anxiety reaction. *Psychosom. Med.*, 1947, **9**, 269.
11. Lantz, B. Some dynamic aspects of success and failure. *Psychol. Monogr.*, 1945, **59**, No. 1 (Whole No. 271).
12. Lazarus, R. S., Deese, J., & Osler, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.
13. Lazarus, R. S., & Eriksen, C. W. The effects of failure stress upon skilled performance. *J. exp. Psychol.*, 1952, **43**, 100–105.
14. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.
15. Pauli, A. E. On a preliminary test for pooling mean squares in the analysis of variance. *Ann. math. Statist.*, 1950, **21**, 539–556.
16. Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
17. Verville, E. The effect of emotional and motivational sets on the perception of incomplete pictures. *J. genet. Psychol.*, 1946, **69**, 133–145.

(Received June 23, 1955)

Journal of Experimental Psychology
Vol. 52, No. 1, 1956

ANTICIPATION OF REWARD AS A FUNCTION OF PARTIAL REINFORCEMENT [1]

HOWARD BRAND, PAUL J. WOODS, AND JAMES M. SAKODA

The University of Connecticut

Brunswik (1) was among the first to consider the effect of varying probability of reward in a two-choice learning task. He considered three hypotheses. The first was concerned with the threshold of probability discrimination, the second with the effect of the probability ratio on learning, and the third with the effect of the probability difference on learning. Five conditions were employed: 100:00, 75:25, 67:33, 100:50, and 50:00. That is, for instance, in the first condition, reward was obtained 100% of the time at one side of a T maze; in the second, the reward was obtained 75% of the time on one side and 25% on the other, and so on. The hypothesis about probability ratio is that if the difference between the two percentages of reward is held constant, for two or more schedules, then the level of learning will increase as the ratio of the percentages increases. The probability difference hypothesis states that if the ratio is held constant, for two or more schedules, then the level of learning will increase as the difference between the percentages increases.

In Brunswik's study, the differences are held constant as the ratio of probability of reward varies in the following manner: 100:50, 75:25, and 50:00 have a difference of 50. The ratio in these varies from 2:1 to 3:1 to infinity. For the constant difference of 50 he found that learning improved as the ratio increased. In a similar way, the probability ratio is held constant. 67:33 and 100:50 have a constant ratio of 2:1; the difference varies from 34 to 50. Likewise, 50:00 and 100:00 schedules have a ratio at infinity; the difference varies from 50 to 100. Here learning improved as the difference increased.

Clearly, these five conditions do not permit an adequate test of the hypotheses because there are insufficient points along the difference dimension and along the ratio dimension. Our main purpose in the present experiment is to rectify this inadequacy, and thereby provide a more satisfactory test of Brunswik's formulation of the influence of probability of reward on two-choice learning.[2] Specifically, we will report on the investigation of the second and third hypotheses described above.

METHOD

General design.—Sixteen reinforcement schedules were used, as shown in Table 1. They were fitted into a double classification factorial design with four levels of ratios and four of differences. Each of the numbers represents the number of reinforcements that were available on each alternative per 100 trials. For example, the reinforcement schedule 40:20 means that the probability of reinforcement on the profitable side was 40 times in 100, and the probability of reinforcement on the unprofitable side was 20 times in 100.

Within a column the ratio is constant as the difference between the two probabilities varies, and within a row the difference is constant as

[1] The present research was conducted under grant M-814 from the National Institute of Mental Health.

[2] This study differs from Brunswik's in that we employed human subjects while Brunswik used animals. The reason for this change is that this study is part of a larger program on human discrimination learning. We know of no reason why these same predictions should not hold for humans as well as animals.

the ratio varies. Each of the 16 reinforcement schedules was assigned randomly to our *Ss*. Since there were five replications for each condition, the number of *Ss* was 80. The *Ss* were run on 120 acquisition and 60 extinction trials.

Subjects.—The *Ss* were 80 undergraduate volunteers from courses in psychology at The University of Connecticut. There were 55 females and 25 males.

Apparatus.—The apparatus consisted of two levers (microswitches) at table height in front of *S* and a vertical panel containing a ground-glass screen at eye level and a green signal light. Red and white bulbs were enclosed in a box behind the vertical panel and flashed on the ground-glass screen when *S* pressed the levers. The levers were controlled independently by *E* through a switching set-up so that each lever could give either color. By instruction, the red light was selected as the reward. The reinforcement was contingent upon *S*'s choice of a lever following a given probability schedule.

Switches on *E*'s control panel enabled him to set up one of four possible situations before each response of *S*. That is to say, the apparatus could be set to give red on the left lever and white on the right lever, or the reverse of this, or red on both levers, or white on both levers. When *E* had the switches set, he flashed the green signal light on *S*'s panel, and *S* pressed a lever. A light on *E*'s control panel told him which lever *S* had pressed. We therefore knew the probability of a red light on each side according to the predetermined probability schedule, the lever that *S* pressed on each trial, and whether or not reward was obtained on each trial.

Instructions.—After *S* was seated comfortably in front of the apparatus, the following instructions were given:

"The two levers in front of you are merely switches which flash either red or white lights on the ground-glass screen. Your task is to try to get as many red lights as you can. But this lever (pressing the one set to give red) will not always be red; sometimes the other lever will be red. The red light will be manipulated back and forth from trial to trial according to a scheme or plan. On a particular trial, then, one lever will be correct and one incorrect. If you press the correct lever, a red light will flash and if you press the incorrect lever, a white light will flash. Your task then is to try to figure out the scheme or plan so that on each trial you can press the one lever that will give you a red light.[3]

"When the green signal light flashes, you are to press one lever and one lever only. Then wait—the green signal light will flash again—this will be the next trial—you then press the lever that you think will be red this time. Remember, each trial you press only one lever, and you press the lever that you think will give you a red light. We can't stop once we have started, so do you have any questions now?"

If there were any questions, *E* paraphrased the relevant part of the instruction, or refused to give additional information until the end of the experiment. Occasionally an *S* would interrupt in the middle of the experiment. When this occurred *E* would say, "Do the best you can."

TABLE 1

SCHEDULES OF REINFORCEMENT CLASSIFIED BY RATIOS AND DIFFERENCES

Differences	Ratios			
	2:1	3:1	6:1	11:1
50	100:50	75:25	60:10	55:5
40	80:40	60:20	48:8	44:4
30	60:30	45:15	36:6	33:3
20	40:20	30:10	24:4	22:2

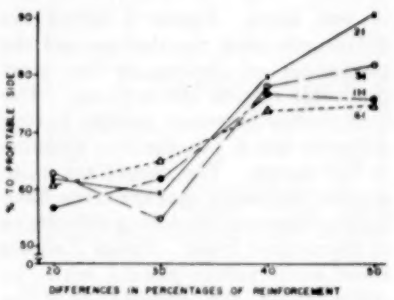

FIG. 1. Percentage of choices to profitable side as a function of probability difference.

RESULTS AND DISCUSSION

The percentage of choices to the most profitable side were plotted for blocks of 10 trials for the acquisition and extinction record of each *S*. A mean value was obtained for the five *Ss* in each cell (i.e., under each schedule of reinforcement) for the last four blocks of acquisition in order to obtain a measure of the asymptotic level. These values may be plotted

[3] Following the convention set up by Estes and Straughan (4).

in two ways. Figure 1 shows the differences along the abscissa and the percentage of choices to the most profitable side on the ordinate. The four curves represent groups having different ratios, i.e., the four columns in the design. The curves are quite similar, indicating that the ratios have little influence in producing differences in asymptotic level. Figure 2 shows these same values plotted with the ratios along the abscissa and the percentage of choices to the most profitable side on the ordinate. The four curves represent groups with varying differences, i.e., the four rows in the design. Here the curves are quite different, indicating that the differences influence the asymptotic level.

A double classification factorial design analysis of variance was also executed on these data. The results are shown in Table 2. Here the findings on the figures are borne out. The only significant variance is that due to the differences in frequency of reinforcement ($P < .01$).

An analysis of the effects of differences and ratios on rate of learning was also undertaken. The rate of learning was measured by linear regression lines fitted to the first five trial blocks. The assumption of linearity does not seem unwarranted as evidenced by the empirical curves plotted in Fig. 3. The analysis performed with the regression data was an analysis of the difference between regressions as outlined by Snedecor (**5**). First of all a test was made of the significance of the average regression to determine whether, in general, the rate of learning was significant or not. Following this the analysis consisted of calculating deviations from average regression and deviations from individual regression and testing the mean square for the difference between the two against an error term. The error term in each case is the mean square of deviations from individual regression. The individual and average regression coefficients are shown in Table 3. The average regression for the table as a whole is significant at the .01 level. The regression coefficients for the four rows are signifi-

TABLE 2

ANALYSIS OF VARIANCE OF ASYMPTOTIC LEVELS

Source	*df*	Mean Square	*F*
1. Differences	3	2356.88	19.59*
2. Ratios	3	95.62	0.79
3. 1 × 2	9	110.14	0.92
4. Within	64	120.31	
5. Total	79		

* $P < .001$.

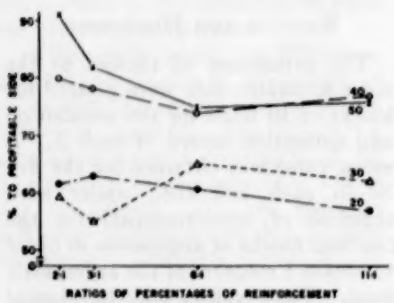

FIG. 2. Percentage of choices to profitable side as a function of probability ratio.

TABLE 3

REGRESSION COEFFICIENTS FOR FIRST FIVE BLOCKS OF ACQUISITION TRIALS

Differences	Ratios				
	2:1	3:1	6:1	11:1	Mean
50	8.2	2.6	2.0	3.6	4.10
40	1.0	1.4	6.4	6.4	3.80
30	1.0	0	2.4	1.8	1.30
20	0	0	−0.8	−0.4	−0.30
Mean	2.55	1.00	2.50	2.85	2.22

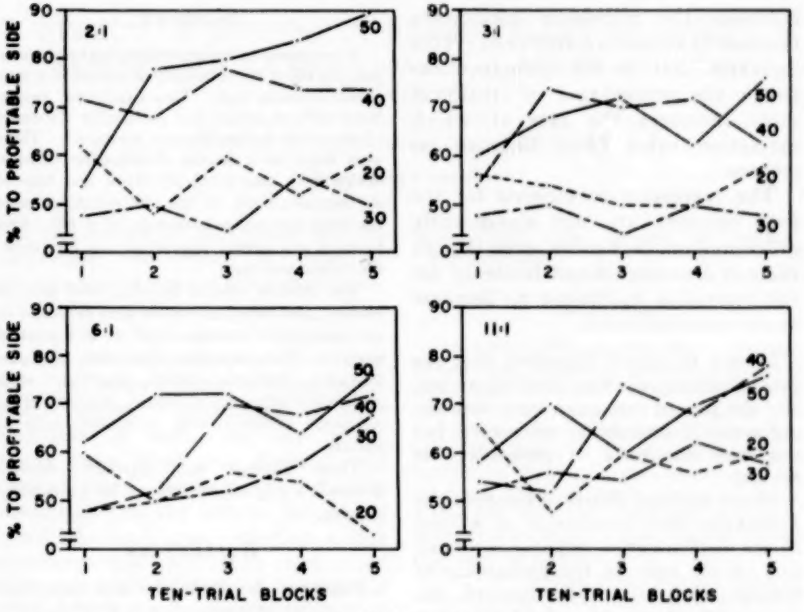

FIG. 3. Percentage of choices to profitable side as a function of trials (in 10-trial blocks).

cantly different from each other ($P < .01$). An inspection of Table 3 shows that as the difference increases the regression coefficients of these lines increase. This indicates that as the difference between the percentages of reinforcement increases the rate of learning likewise increases.

The regression coefficients for the four columns are not significantly different from each other. An inspection of Table 3 shows that as the ratios vary the regression coefficients do not vary in any consistent manner. This indicates that a change in the ratios of the percentages of reinforcement does not alter in any consistent manner the rate at which learning will take place.

An analysis similar to that just outlined was performed on the extinction data. Regression coefficients were calculated for the last block of acquisition and the first three blocks of extinction trials. The individual and average regression coefficients are shown in Table 4. The average of the regression for the table as a whole is significant ($P < .01$). The regression coefficients for the four rows are also significantly different from each other ($P < .01$). An inspection of Table 4 shows that as the difference

TABLE 4

REGRESSION COEFFICIENTS FOR LAST BLOCK OF ACQUISITION AND FIRST THREE BLOCKS OF EXTINCTION TRIALS

Differences	Ratios				
	2:1	3:1	6:1	11:1	Mean
50	−10.2	−15.2	−9.4	−6.6	−10.35
40	−11.4	−7.0	−5.0	−5.0	−7.10
30	−6.0	2.6	−2.6	−2.4	−2.10
20	1.4	0	−4.4	3.6	0.15
Mean	−6.55	−4.90	−5.35	−2.60	−4.85

increases the regression coefficients increase in a negative direction. This indicates that as the difference between the percentages of reinforcement increases the rate at which extinction takes place likewise increases.

The regression coefficients for the four columns are not significantly different from each other, even though there is a nonsignificant tendency for the regression coefficient to decrease as the ratio increases.

It may be stated, therefore, that the rate of learning, the final level of learning, and the rate of extinction vary with the *difference* of probability of reward but not with the *ratio* of probability of reward.

These findings clearly disconfirm the hypothesis that acquisition of a two-choice discrimination response is a function of the ratio of the probability of reward when the difference of the probability of reward is held constant. We do, however, obtain the result that acquisition and extinction of the response is dependent upon the difference in the probability of reward. Here, Brunswik (**1**, p. 185) is confirmed. The absence of any interaction between difference of probability and ratio of probability further indicates the relationship is unequivocal.

For the range of differences in reward sampled here, we have found no indication of a threshold value for discrimination of probability of reward. If a threshold had been obtained, then there would be an interaction between ratio of reward and difference of reward. This did not occur. Learning and extinction in this experiment were always a direct function of the increasing difference of reward on one side of the discrimination versus reward on the other side.[4]

[4] These results may be interpreted in terms of a statistical learning model (**3**). This model, however, as Brunswik (**2**) has already pointed out is, in principle, different from what he would propose as the role of probability in learning because the experimental designs lack *representativeness*. If *representativeness* is an empirical issue, however, it would be expected that where it has been built into an experiment, as it was in this study, there should be confirmation of Brunswik's theoretical position. Unhappily, the contrary was found to be the case.

Summary

The purpose of the experiment was to investigate the effect of probability of reward in a two-choice learning task. Two conditions, probability ratio of reward and probability difference of reward on the two choices, were used. These were fitted to a double classification factorial design with four levels of ratios and four of differences. Each of the 16 schedules were randomly assigned to human *Ss* ($N = 80$). One hundred and twenty acquisition and 60 extinction trials were run.

The data on rate of learning, final level of learning, and rate of extinction were subjected to an analysis of variance and to a regression analysis. Both analyses show that the performance measures varied positively with probability difference of reward, but showed no consistent variation with probability ratio of reward.

These findings only partially confirm Brunswik's original hypotheses on probability learning.

References

1. Brunswik, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–197.
2. Brunswik, E. Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.*, 1955, **62**, 193–218.
3. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
4. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
5. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Iowa State College, 1946.

(Received July 20, 1955)

Journal of Experimental Psychology
Vol. 52, No. 1, 1956

EFFECT OF DIFFERENT PERCENTAGES OF MONEY REWARD ON EXTINCTION OF A LEVER-PULLING RESPONSE [1]

DONALD J. LEWIS AND CARL P. DUNCAN

Northwestern University

The purpose of this study was to determine the effect of different percentages of reward upon extinction of a lever-pulling response. The study differs from most others in the area of partial reinforcement in three principal ways. First, a larger number of different percentages of reinforcement is used than have previously been employed. Second, the reward used (money) was highly meaningful to the human *Ss*. And, third, the task used—playing a slot machine—was one that had considerable intrinsic interest for the *Ss*.

One of the more reliable "laws" of experimental psychology is that concerning the effects of partial reinforcement which holds that a partially reinforced response is more resistant to extinction than a continuously reinforced one. This law has been demonstrated to hold for a considerable number of stimulus situations and response systems (**3**). Attempts to determine the extent of application of the law of partial reinforcement are important for at least two reasons: (*a*) theorists (**1**) have made important extensions of behavior-theory principles, admitting that the generality of these principles needs to be determined experimentally, and (*b*) studying stimulus situations and response systems different from those commonly used may well yield information about the nature of partial reinforcement.

[1] This study was supported in part by a grant from the National Science Foundation. The authors wish to thank S. A. Kendrick, C. Meyer, J. Mintz, and E. Lindquist, for their assistance with this experiment.

METHOD

Apparatus.—The apparatus was a standard 25-cent slot machine ("one-armed bandit") which had undergone modification for the purposes of this research. The machine was played with chromium-plated steel discs which, upon being placed by *S* in the coin slot, eventually found their way to the cash drawer of the machine. The cash drawer contained almost a thousand discs, and constituted the supply reservoir upon which *S* drew when he needed more discs for his play.

The *S* was paid off from a separate supply of discs which were stored in a tube that led directly to the payoff tray on the front of the machine. Release of a disc from the tube into the payoff tray was accomplished by activation of a trigger attached to the tube. The trigger was tripped by brads held in holes drilled in the perimeter of a wheel which was mounted inside the machine. The *E* determined beforehand which plays would result in payoffs simply by inserting brads into holes in the wheel.

A counter mounted inside the machine cumulated the total number of plays for each *S*.

Since the machine was played with discs, rather than quarters, and since *S* was never rewarded with more than one disc even on a payoff play, the 25¢ symbol and the plaque describing the original payoff system, both of which appeared on the front of the machine, were covered with masking tape. The pictures of the various fruits and other objects, which revolved behind the glass window with each play, were unchanged. Except for the masking tape and the use of discs, the machine looked and sounded like an ordinary slot machine.

Design.—A total of seven groups of 50 *Ss* each were run, each group having an "acquisition" and an "extinction" series. The acquisition series, for each group, consisted of the first eight plays. After the acquisition series the extinction series began; no *S* ever won after the eighth play.

It was thought that perhaps *pattern* of rein-

forcement might be important, especially with low percentages where pattern can vary a good deal within a percentage. So a pilot study on patterns was carried out, using 25% reinforcement. Four groups, 25 *Ss* each, were paid off on the following patterns: Trials 1 and 8, 3 and 8, 5 and 8, and 7 and 8. The mean log plays to extinction for these groups did not differ significantly; F was less than one. Therefore, the pattern for the other partial reinforcement groups was chosen rather arbitrarily according to the criteria described below.

Payoff trials were selected to be irregular insofar as possible within any percentage group, and so that if any payoffs were to be given at all, one would be on Trial 8. This made it meaningful to talk about an 8-trial acquisition series. The plays that resulted in a reward are shown by Table 1.

TABLE 1

THE PATTERN OF PAYOFFS (X's)

Group	Trials							
	1	2	3	4	5	6	7	8
100%	X	X	X	X	X	X	X	X
75%	X	0	X	X	X	0	X	X
50%	X	0	X	0	0	X	0	X
37½%	0	0	X	0	X	0	0	X
25%	0	0	X	0	0	0	0	X
12½%	0	0	0	0	0	0	0	X
0%	0	0	0	0	0	0	0	0

Assignment of Ss.—The *Ss* were 350 men and women from the introductory psychology classes at Northwestern University. The experiment was run in two phases. Assignment of *Ss* for Phase 1 was as follows: The 100% and 50% groups first received a complement of 30 *Ss*. Then the 25% and 0% groups each received 30 *Ss*, and finally 30 *Ss* were assigned to the 12½% 37½%, and 75% groups. Phase 2 of the experiment then began. Twenty additional *Ss* were assigned to each of the seven groups in turn as *Ss* came to the laboratory. The *Ss* for the total experiment were drawn from six different academic quarters (and from all seasons of the year).

Because of the method of assignment used during Phase 1—a method made necessary by practical considerations—there was a possibility that group differences were caused by assignment rather than by the experimental variable. To determine whether this happened, data from the two phases were analyzed separately.

Instructions.—The instructions, which were spoken to S from memory, were as follows:

"Have you played a slot machine before? [If S says, "yes"] Then you know there is nothing complicated about it. [If S says, "no"] Well, it's very simple; I'll show you. You play the machine with these discs out of the drawer back here. [About 15 discs are taken from the cash drawer while S watches.] Whenever you want more discs to play with, just take them out of this drawer; then push the drawer back in. [E pushes drawer back.] You play this machine in the usual way. Just put a disc in the slot and pull the lever all the way down. The idea is to win as many of these discs as you can. The ones you win will fall into this tray here, and later I'll pay you a nickel for each one you win. Don't use the ones you win to play with. Whenever you want more to play with, just take them out of the drawer. You can play as long as you want. Just keep taking discs out of the drawer and play until you decide to quit. Then bring over what you have won and cash them in. I won't answer any more questions now or while you are playing, but when you quit, bring over the discs you have won, cash them in, and we will talk about it at that time. One other thing. You can't influence any slot machine by how hard you pull the lever, so just give it a normal pull all the way down. Remember, you can play as long as you want. Okey, begin when you are ready."

Procedure.—All *Ss* were run individually. Preliminary experimentation indicated that a 30-min. experimental period was sufficiently long to allow S to play as long as he wanted. Only two *Ss* would have exceeded this time limit if allowed to continue; their data were discarded.

While S played E sat reading at a table in the same room. When S was through playing, he was given a nickel for each disc that he had won. He was then told that he had participated in a study on incidental learning, and was asked if he could remember what combinations of fruit pictures had paid off, which ones had not, which ones came up most often, etc. He was then asked not to tell anyone what the purpose of the study was. It was hoped that this would prevent the true nature of the experiment from being known by *Ss* before and after they played. There was no evidence to indicate that this was not the case.

RESULTS AND DISCUSSION

The raw score for each S was simply the total number of plays (including the eight acquisition plays). Because the distributions of raw scores for all groups were positively skewed, the

scores were transformed into logs which resulted in a satisfactory normalization.

Since the experiment was run in two phases, as described earlier, it was necessary to determine if there were phase differences that affected the experimental group differences. The F less than 1 (see Table 2) attributable to the phases indicates that the assignment of Ss did not result in significant differences, and the interaction F, again less than 1, indicates that the phase assignment did not interact significantly with the major variable.

In Table 2 the F of 4.41, which is significant beyond the .01 level, indicates that the percentage of reinforcement had a very significant effect upon total number of plays. This effect is shown graphically in Fig. 1. It seems clear that the persistence of the lever-pulling response is an inverse function of the percentage of reward. That partial reinforcement is effective in increasing the resistance to extinction of the response as compared to continuous reinforcement is indicated by the t of 3.72 for the comparison between the 100% group and the pooled partial reinforcement groups. It also appears that no reinforcement at all increases the total number of plays as compared to continuous reinforcement ($t = 6.12$).

Grant and Schipper (2) determined the effectiveness of 0%, 25%, 50%, 75%, and 100% reinforcement on the resistance to extinction of a conditioned eyelid response. They found that resistance to extinction was greatest in the 50% and 75% groups, falling off somewhat in the 25% and 100% groups. The 0% group showed very little initial conditioning and consequently no resistance to extinction. Grant and Schipper interpreted their results in terms of two processes. The first had to do with the habit strength at the beginning of extinction. If the habit is weak it will extinguish quickly, even though it was acquired under a partial reinforcement schedule. Grant and Schipper stated that the 25% reinforcement schedule resulted in inferior resistance to extinction because it in all probability gives barely sufficient reinforcement for the response to appear consistently during acquisition. The second process they describe as the ease of discriminating the acquisition series from the extinction series, and they believe that this discrimination is a function of the similarity between acquisition and extinction. Putting these two processes together, the greatest resistance to extinction should occur both where response strength is relatively large and discrimination difficult. This maximum occurred somewhere between 50% and 75% reinforcement in the Grant and Schipper situation.

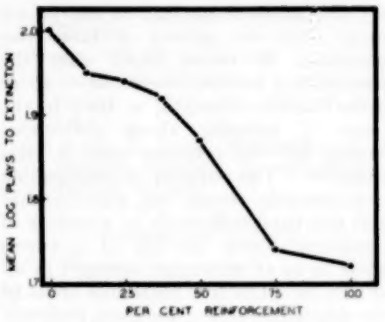

FIG. 1. Mean log trials to extinction as a function of the percentage of trials reinforced during acquisition.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE

Source	*df*	Mean Square	*F*
% Reinforcement	6	.6178	4.41
Phase	1	.0387	
Interaction	6	.1046	
Within	337	.1401	

Our results can perhaps be interpreted using only the second of these two processes. It seems likely that the schedules of reinforcement had no effect upon response strength, at least in the sense of bringing about differential habits, for the response used in this situation. The response of lever-pulling is a relatively simple one, and one that gets practiced sufficiently in a variety of situations during the life of a young adult to be at maximum strength in all of our *S*s. The few additional trials of the acquisition series, therefore, probably did not increase the habit strength of any of the *S*s. These acquisition trials did have an effect, however, in setting the acquisition series apart from the extinction series, the second of Grant and Schipper's two processes.

Because continuous reinforcement is most unlike continuous nonreinforcement, the discrimination between acquisition and extinction should be easiest for the 100% *S*s and they should quit first when extinction starts. Ease of discrimination should be next greatest for the 75% group, then for the 50% group, and so on through the 12½% group. Using only the ease of discrimination process, a general inverse relation should result between percentage reinforcement and resistance to extinction, such as shown in Fig. 1.

The 0% group of the present experiment is somewhat unusual in that it does not represent a response of weak or no habit strength as it does in the Grant and Schipper experiment and in most experiments where it has been used. The discrimination process, however, seems helpful in explaining the results from this group also. Because continuous nonreinforcement is identical with further continuous nonreinforcement, the difference between the first eight plays and the succeeding plays is negligible, making an extinction series difficult to discriminate.

Lewis (**4**), using 7-yr.-old boys as *S*s and button-pushing response, found that 50% and 60% reinforcement resulted in superior resistance to extinction over 100%, but that the partial percentages also resulted in superior resistance to extinction over 0%. The difference between 0% and 100% was in favor of the 0% group but was not significant. A graph of his results yields a curve much like that given by Grant and Schipper instead of the nearly linear trend of the present experiment. Perhaps for such young *S*s, with restricted learning opportunities, we must also utilize the first of the two processes. Assuming the button-pushing response to be below maximum strength in the young boys, the response would be increased as a function of the number of reinforcements. The effects of reinforcement together with the process of discrimination could result in a maximum at about 50% to 60% reinforcement.

It could also be argued that the schedule of reinforcement may affect the motivation of *S*s. If one considers the general motivation of *S*s in this experiment to be the desire to win, then this motivation would be reduced most by 100% reinforcement, less by 50% reinforcement and least by 0% reinforcement, again giving us a trend between percentage of reinforcement and number of plays to extinction.

Summary

The purpose of this study was to determine the effect of seven different percentages of reward upon extinction of a lever-pulling response. An ordinary slot machine was modified so that *E* could determine both the total number of payoffs (rewards) and the exact trial on which payoff would be made. Seven groups of fifty human *S*s each were run under different percentages—100%, 75%, 50%, 37½%, 25%, 12½%, and 0%—of reward. The acquisition series was the first eight plays, after which no *S* ever won again. The response measure was the total number of plays prior to *S*'s decision to quit. There was a general inverse relation between percentage of reinforcement and number of plays to quitting. The ease of discriminating between acquisition and "extinc-

tion" was believed to play an important role in determining the inverse relation.

References

1. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
2. Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43,** 313–320.
3. Jenkins, W. O., & Stanley, J. C., Jr. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47,** 193–234.
4. Lewis, D. J. Partial reinforcement in a gambling situation. *J. exp. Psychol*, 1952, **43,** 447–450.

(Received June 13, 1955)

Journal of Experimental Psychology
Vol. 52, No. 1, 1956

AN ANALYSIS OF THE EFFECT OF NONREINFORCED TRIALS IN TERMS OF STATISTICAL LEARNING THEORY

RICHARD C. ATKINSON[1]

Indiana University

It is the purpose of this study to investigate the effect of nonreinforced trials in a Humphreys-type "verbal conditioning" situation. Before specifying the problem, however, a synopsis of both the situation and the theoretical orientation to be employed will be presented by referring to an article by Estes and Straughan.

We consider a situation in which each trial begins with presentation of a signal, or CS; following the signal, one or the other of two reinforcing stimuli, E_1 or E_2, occurs, the probability of E_1 and E_2 during a given series being π and $1 - \pi$, respectively. The behaviors available to S are categorized into two classes, A_1 and A_2, by experimental criteria. In the verbal conditioning situation, A_1 is a prediction that E_1 will occur, and A_2 a prediction that E_2 will occur on the given trial. We assume that the CS determines a population, S_e, of stimulus elements which is sampled by S on each trial, the proportion θ of the elements in this population constituting the effective sample on any one trial. The dependence of S's responses upon the stimulating situation is expressed in the theory by defining a conditional relationship such that each element in S_e is conditioned to (tends to evoke) either A_1 or A_2. In order to interpret the formal model in terms of a verbal conditioning experiment, we assume that when an E_1 occurs it evokes from S a response belonging to class A_1, i.e., one which is compatible with the response of predicting E_1 but which interferes with the response of prediction E_2, and that when an E_2 occurs it evokes a response of class A_2. Then on a trial on which E_1 occurs we expect on the basis of association principles that all elements sampled from S_e on the trial will become conditioned to A_1, while on an E_2 trial the sample will be conditioned to A_2. Now if successive trials are sufficiently discrete so that samples from S_e are statistically independent, the probability of an A_1 after Trial n, abbreviated $p(n)$, is defined in the model as the proportion of elements in S_e that are conditioned to A_1, and similarly for the probability of an A_2, $[1 - p(n)]$ (**2**, p. 225).

An interesting problem arises when one modifies this situation by introducing trials on which neither E_1 nor E_2 occurs, that is, trials on which S is given no information as to whether or not his predictive response (A_1 or A_2) was correct. What prediction would one venture concerning the effect of such nonreinforced trials on the probability of an A_1 or A_2 response on subsequent trials? This paper proposes and examines the following three hypotheses which are stated in terms of statistical association concepts (**1**).

Hypothesis I, (H-I).—Nonreinforced trials have no effect on the hypothetical connection between stimulus elements and response classes. Elements sampled from the S_e population on such trials are returned to the population at the end

[1] The author wishes to thank Professor W. K. Estes for advice and assistance in carrying out this research.

of the trial with their conditional status unchanged. That is to say, if an element is sampled on a nonreinforced trial, it will remain connected to the same response class that it was connected to before the start of the trial.

Hypothesis II, (H-II).—In the absence of the E_1 and E_2 events, S's overt predictive A_1 or A_2 response is viewed as a terminal response for the trial, and it is posited that elements sampled on such nonreinforced trials are conditioned to the A_1 or A_2 response emitted on the trial. This notion is similar to Guthrie's assumption (**3**) that the last response emitted in the presence of a stimulus pattern is the one which will tend to be produced when the stimulus pattern recurs.

Hypothesis III, (H-III).—Elements sampled on a nonreinforced trial are connected to the A_1 response class with probability σ, and to the A_2 response class with probability $1 - \sigma$. We shall only consider the case where $\sigma = 1 - \sigma = .5$. Here we assume that, in the absence of information as to which response was correct, there exists an equal likelihood that sampled elements will be connected to either response class.

In order to evaluate these hypotheses we shall use an approach similar to that employed by Estes and Straughan (**2**). For any schedule of E_1 and E_2 trials they were able, by means of statistical learning theory, to predict asymptotic response probability. Similarly, for any schedule of E_1, E_2, and nonreinforced trials we can predict asymptotic response probability under the conditions of each of the hypotheses stated above. The problem then of defining the effect of nonreinforced trials on behavior can be formulated in terms of the long run effect of schedules of reinforced and nonreinforced trials on asymptotic response probability.

There is, however, the problem that not all schedules of E_1, E_2, and nonreinforced trials will yield differential predictions for the three hypotheses. For example, in the situation employed by Estes and Straughan (**2**), it can be shown that the inclusion of nonreinforced trials with any fixed probability in the schedule of E_1's and E_2's would yield identical asymptote and slope predictions (*a*) for H-I and H-II for all values of π and (*b*) for H-I, H-II, and H-III for $\pi = .5$.

Consequently, a schedule of events is required which will yield differential predictions for the three hypotheses. We have selected the following schedule which, with the substitution of appropriate experimental parameter values, meets this requirement:

1. If an A_1 occurs on Trial n, there is: (*a*) a probability $\beta\pi$ that an E_1 will occur, (*b*) a probability $\beta(1 - \pi)$ that an E_2 will occur, (*c*) a probability $(1 - \beta)$ that neither an E_1 nor E_2 will occur.

2. If an A_2 occurs on Trial n, there is: (*a*) a probability π that an E_1 will occur, (*b*) a probability $(1 - \pi)$ that an E_2 will occur.

By methods of finite calculus we can describe the expected asymptotic probability of an A_1 response, denoted by $\overline{p(\infty)}$, for the above schedule under the conditions of each hypothesis.

By H-I, $\overline{p(\infty)}$ is bounded by

$$\pi \quad \text{and} \quad \frac{\pi}{\pi + \beta(1 - \pi)}. \tag{1}$$

By H-II,

$$\overline{p(\infty)} = \frac{\pi}{1 - (1 - \beta)(1 - \pi)}. \tag{2}$$

By H-III,

$$\overline{p(\infty)} = \frac{\pi}{1 - (1 - \beta)(.5 - \pi)}. \tag{3}$$

The derivations of these results are given in the next section.

Mathematical Formulation

For the schedule specified above, the probability of an A_1 response after Trial n, $p(n + 1)$, is:

By H-I,

$$p(n+1) = (1-\theta)p(n) + \theta\{[1-p(n)] \times [\pi(1) + (1-\pi)(0)] + p(n)\beta[\pi(1) + (1-\pi)(0)] + p(n)(1-\beta)p(n)\}. \quad (4)$$

The proportion $(1-\theta)$ of the elements is not sampled and remains unchanged. The proportion θ of the elements is sampled and conditioned as follows: (*a*) If an A_2 occurs, with probability $[1-p(n)]$, there is a probability π that all sampled elements will be connected to A_1, and a probability $(1-\pi)$ that the elements will be connected to A_2; (*b*) If an A_1 occurs, with probability $p(n)$, there is a probability $\beta\pi$ that all sampled elements will be conditioned to A_1, a probability $\beta(1-\pi)$ that the elements will be conditioned to A_2, and a probability $(1-\beta)$ that the conditional status of the elements will remain unchanged.

Rewrite Equation 4

$$p(n+1) = (1-\theta)p(n) + \theta[\lambda(n) + p(n)^2(1-\beta)] \quad (5)$$

where

$$\lambda(n) = [1-p(n)]\pi + p(n)\beta\pi. \quad (6)$$

By H-II,

$$p(n+1) = (1-\theta)p(n) + \theta[\lambda(n) + p(n)(1-\beta)]. \quad (7)$$

This equation is isomorphic with Equation 5 except for the final product of the last term. If an A_1 occurs and is not followed by an E_1 or E_2, this joint event having probability $p(n)(1-\beta)$, then all elements will be conditioned to the response emitted on the trial, namely, A_1.

By H-III,

$$p(n+1) = (1-\theta)p(n) + \theta[\lambda(n) + p(n)(1-\beta)(.5)]. \quad (8)$$

This equation is also isomorphic with Equation 5 except for the final product of the last term. If an A_1 event occurs and is not followed by an E_1 or E_2, then there is a probability .5 that sampled elements will be connected to A_1.

Taking the expected value of Equation 7 and solving the resulting equation (4) we obtain

$$\overline{p(n)} = \overline{p(\infty)} - [\overline{p(\infty)} - \overline{p(0)}] \times [1-\theta+\theta(1-\beta)(1-\pi)]^n, \quad (9)$$

where $\overline{p(\infty)}$ is given by Equation 2.

Similarly for Equation 8

$$\overline{p(n)} = \overline{p(\infty)} - [\overline{p(\infty)} - \overline{p(0)}] \times [1-\theta+\theta(1-\beta)(.5-\pi)]^n, \quad (10)$$

where $\overline{p(\infty)}$ is given by Equation 3.

Equation 5 is a second-order difference equation, and no method is known for effecting a solution. However, by taking the expected value and letting $n \to \infty$ we obtain

$$\overline{p(\infty)} = \frac{\pi + (1-\beta)\overline{p(\infty)^2}}{1+\pi(1-\beta)}. \quad (11)$$

An upper and lower bound can now be calculated for $\overline{p(\infty)}$ by letting the variance of $\overline{p(\infty)}$ assume a minimum value of zero and a maximum of $\bar{p}(1-\bar{p})$. When variance is minimal $\bar{p}^2 = \overline{p^2}$, when maximal $\bar{p} = \overline{p^2}$. These substitutions in Equation 11 yield boundary conditions on $\overline{p(\infty)}$ given by Equation 1.

Method

Design.—Two groups, A and B, were necessary for good differentiation between the hypotheses. For Group A the value of π was .8125; for Group B the value of π was .1875. The value of β was .25 for both groups.

Apparatus.—The *E* and *S* sat on opposite sides of a table with a large opaque screen separating them. The apparatus, as viewed from *S*'s side, consisted of two telegraph keys mounted 24 in. apart at right angles to the base of the screen; upon the screen were mounted three milk-glass panel lights. One of these lights, which served as the signal for *S* to guess, was centered between the telegraph keys at *S*'s eye level. Each of the two remaining lights, the reinforcing signals, was mounted directly above one of the telegraph keys. The presentation and duration of the lights were automatically controlled, and on any trial the reinforcing signal could be withheld.

Subjects.—The *S*s, 52 students obtained from beginning courses in psychology during the fall semester of 1954, were assigned at random to experimental groups.

Procedure.—After being seated in front of the screen *S* was read instructions, similar to those reported by Estes and Straughan (2), requesting that he respond to the signal light on each trial by operating the telegraph key corresponding to the reinforcing light that he expected to follow. For each *S*, one of the two reinforcing lights was randomly designated as E_1, and the other as E_2. After the instructions were read, four practice trials were given with the lights presented in the order E_2, E_1, E_1, E_2. After these trials *E* corrected any apparent errors in procedure and reviewed the instructions.

Following the practice trials, 144 trials were given each *S*. These trials were run off in continuous sequence. For each *S* a sequence of E_1, E_2, and nonreinforced trials was generated in accordance with the assigned values of π and β.

On each trial the signal light was lighted for 2.5 sec.; the time between successive signal exposures was 7.5 sec. On reinforced trials the reinforcing light followed the cessation of the signal light by 2 sec. and remained on for 1 sec.

Results and Discussion

Substituting values of π and β in Equations 1, 2, and 3, $\overline{p(\infty)}$ can be described under the conditions of each hypothesis.

Predictions for Group A are

1. By H-I, $\overline{p(\infty)}$ is bounded by .8125 and .9455.
2. By H-II, $\overline{p(\infty)} = .9455$.
3. By H-III, $\overline{p(\infty)} = .6582$.

Predictions for Group B are

1. By H-I, $\overline{p(\infty)}$ is bounded by .1875 and .4800.
2. By H-II, $\overline{p(\infty)} = .4800$.
3. By H-III, $\overline{p(\infty)} = .2449$.

Taking the proportion of A_1 responses over the last two blocks of 16 trials as an estimate of $\overline{p(\infty)}$ we can compare these values with the above set of predictions. By inspection of Fig. 1 we see that for Group A the proportions for the last two trial blocks are .828 and .842, with a mean of .835. To determine how well this value approximated the predictions of H-II and H-III, t tests were run between the predicted value and the proportions computed for individual Ss over the last 32 trials. The σ_m was .017. The difference between the predicted value and the mean of the observed values was .110 for H-II and .175 for H-III; both differences are significant at the .05 level. By inspection of Fig. 1 we see that for Group B the proportions for the last two trial blocks are .315 and .337 with a mean of .326. The σ_m was .021. The differences between predicted and observed values were .154 and .081 for H-II and H-III, respectively; both are significant at the .05 level.

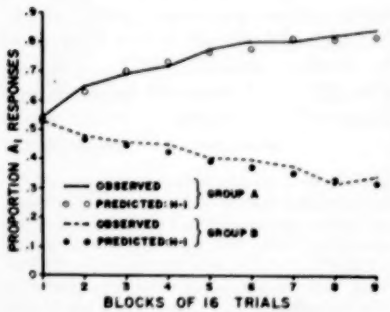

Fig. 1. Empirical and theoretical curves representing mean proportion of A_1 responses per 16-trial blocks.

In view of these results, it is clear that neither H-II nor H-III accounts for the obtained data. What then can be said concerning H-I? First of all, the proportions computed over the last 32 trials for both groups fall within the bounds for $\overline{p(\infty)}$ specified by H-I. However, more important is the fact that we can generate curves by H-I which closely approximate the observed results. By employing a Monte Carlo method (**5**) 10 individual learning curves were generated under the conditions of Equation 4, five using the experimental parameter values of Group A, and five using those of Group B. The θ and $p(0)$ values employed for Group A were .063 and .480; respective values for Group B were .010 and .550. The averages by blocks of 16 trials for each group of five curves are presented in Fig. 1. As can be seen, there is close agreement between these values and the observed values.

We conclude that, of the three hypotheses presented, only H-I can account for the findings of this study.

Summary

This study employs a modified Estes-Straughan "verbal conditioning" situation. Each trial consists in the presentation of a signal followed by an E_1 or E_2 reinforcing light; S operates an appropriate key to indicate his prediction as to which light will appear on each trial. The present study deals with the effect of introducing trials on which neither reinforcing light is presented, that is, trials on which S is given no information as to which of the two alternative predictions is correct.

Three hypotheses are developed, in terms of statistical learning theory, concerning the effect of such trials. Since some schedules of E_1, E_2, and nonreinforced trials provide identical predictions from the three hypotheses, it was necessary to employ a schedule which yielded differential prediction. The one selected was such that on any trial (*a*) if S predicted E_1 then E_1, E_2, or no reinforcement would occur with probability $\beta\pi, \beta(1-\pi)$ and $1-\beta$, respectively, while (*b*) if S predicted E_2 either E_1 or E_2 would occur with probability π and $1-\pi$.

The set of quantitative predictions generated by each hypothesis was checked against experimental results. It was concluded that the hypothesis requiring zero change in response probability on trials on which no information is given best accounts for the findings.

References

1. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
2. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
3. Guthrie, E. R. Psychological facts and psychological theory. *Psychol. Bull.*, 1946, **43**, 1–20.
4. Jordan, C. *Calculus of finite differences.* New York: Chelsea, 1950.
5. McCracken, D. D. The Monte Carlo method. *Scientific Amer.*, May 1955, 90–96.

(Received June 30, 1955)

Journal of Experimental Psychology
Vol. 52, No. 1, 1956

A METHODOLOGICAL STUDY OF THE PREPARATION OF CONNECTED VERBAL STIMULI FOR QUANTITATIVE MEMORY EXPERIMENTS [1]

EUGENE E. LEVITT

Institute for Juvenile Research, Chicago

The problem of subdividing connected verbal material for quantitative scoring of recall occurs in the construction of memory scales, in laboratory studies of rumor transmission, and in investigations of the effects of attitude, frame of reference, and other variables on memory. The manner in which the material is divided may have significant effects on experimental results, yet few researchers have paid much attention to it. Not infrequently the only reference to preparation for scoring in an experimental report is the notation that the material was divided into "items," or "ideas." This technique is the one which is commonly used in breaking up prose for quantitative analysis. That is to say, the stimulus passage is divided into a number of "meaning units." A meaning unit is a linked set of words, usually less than a sentence, which *E* regards as an ideational entity. The procedure was probably first applied by Henderson (**3**) over fifty years ago, and has since been used widely in investigations of memory. The following subtest of the Wechsler Memory Scale (**12**) will serve as an illustration. The units are separated by diagonals.

Anna Thompson/ of South/ Boston/ employed/ as a scrub woman/ in an office building/ reported/ at the City Hall/ Station/ that she had been held up/ on State Street/ the night before/ and robbed/ of fifteen dollars/. She had four/ little children/ the rent/ was due/ and they had not eaten/ for two days/. The officers/ touched by the woman's story/ made up a purse/ for her/.

The procedure which is now employed is the classic method of retained members, in which it is assumed that the stimulus cannot be recalled totally by *S*, and his recall score will be the number of units which he can reproduce to the satisfaction of *E*. There are a number of different ways of determining adequate recall and of assigning partial credit, such as those of Henderson (**3**), Lyon (**8**), and Clark (**2**), among others. The present investigation, however, is concerned only with division of the stimulus.

It is evident that any method for creating meaning units or idea groups will be more or less arbitrary. There have been no attempts to define idea groups empirically, that is, based on the performances of a group of *S*s, and there is considerable doubt that such a task could be accomplished satisfactorily. The fact that the method will be entirely at *E*'s whim makes it all the more unfortunate that few seem concerned with the methodology. Suppose, for instance, that a given *S* recalls 6 idea groups of the 24 in Wechsler's passage. His recall score is then 25%. His performance included the phrase "of ten dollars,"

[1] This paper is based in part on a section of a dissertation submitted in partial fulfillment of the degree of Doctor of Philosophy in the Department of Psychology, Columbia University, in 1952. The author is indebted to Drs. Otto Klineberg and Henry E. Garrett for their direction of the dissertation. The experimental sections were done while the author was at the Child Welfare Research Station, University of Iowa.

instead of the correct "of fifteen dollars." Hence he received no credit for this idea group. Now if we had divided the phrase originally into two idea groups, thus: / of fifteen/ dollars/, *S* would have been credited with one of two idea groups where he received zero credit before. His recall score is now 7/25 = 28%. We have increased this *S*'s score by fully 12% of his original score with one simple change of a meaning unit.

The full significance of this alteration can be seen if we realize the multiplicity of ways in which a prose passage can be divided for scoring. The number of possible arrangements can be computed by means of the following formula if the maximum number of words per idea group is known:[2]

$$A = xn - \frac{x(x-1)}{2}$$

in which n is the number of words in the passage and x is the maximum number of words per IG. For Wechsler's passage $n = 68$ and $x = 6$. The number of possible arrangements is 393. This total is somewhat misleading since any prose contains a certain number of minor parts of speech like the preposition "of" which could logically be grouped with only one set of words. If we deal only with nouns, pronouns, verbs, and adjectives, for Wechsler's passage, $n = 45$, $x = 5$, and $A = 215$. Even if we cut down the maximum words per IG to 3 there will still be 132 possible arrangements. No matter what kind of allowances are made there will nearly always be a considerable number of arrangements even for a relatively short piece of prose. When the stimuli contain several hundred words as in the studies of Clark (**2**), and Levine and Murphy (**5**), the number of arrangements approaches or exceeds a thousand.

A particular arrangement can be described in terms of the mean number of words per IG and the *SD* of the distribution of words per IG. These data are shown in Table 1 for six prose passages taken from three studies of functional memory (**2, 5, 11**).

TABLE 1

MEANS AND *SD*'S OF WORDS PER IDEA GROUP IN THREE STUDIES

Study	Passage	Mean	*SD*	No. Words
Taft (**11**)	I	3.75	2.26	120
Levine-Murphy (**5**)	I	4.20	1.90	294
	II	4.89	2.86	313
Clark (**2**)	I	3.87	1.79	232
	II	3.44	1.33	172
	III	3.38	.98	169

Inspection of Table 1 suggests that there are significant differences in variance from passage to passage. This is confirmed by a Bartlett's test which yields a χ^2 of 52.08, $P < .001$ for 5 *df*. This does not preclude a variance analysis of the means if we assume that the shapes of the distributions do not differ markedly. An *F* ratio among the means is 6.14, which is significant beyond the .001 level, and hence may be accepted as indicating significance at the .01 level (cf. **7**).

We may conclude that different *E*s have used varying kinds of unit groupings in experimental memory studies. There is a tendency for the mean number of words per IG to fluctuate directly with the number of words in the passage. These differences may account, at least in part, for conflicting experimental reports, especially reports of the percentage of recall for a given sample. They may have confounded results in studies in which the ability of a homogeneous group to recall material which was favorable (in the broadest sense of the word) to the group was contrasted with its ability to recall an unfavorable or neutral passage. Many of the investigations of functional mnemonics, such as those of Alper and Korchin (**1**),

[2] The author wishes to express his gratitude to Dr. Robert Plutchik of Hofstra College for devising this formula.

TABLE 2

NUMBER OF IDEA GROUPS PER PASSAGE FOR FOUR PASSAGES, AS DETERMINED BY 31 *Es*

Passage	Words	Idea Groups		
		Mean	*SD*	Range
A	58	10.61	6.38	3–30
B	46	9.03	5.11	2–18
C	67	11.97	6.09	2–24
D	53	10.29	6.06	2–21
Mean	56.25	10.48	5.58	2.5–22.5

Clark (2), Levine and Murphy (5), Menzies (9), and Taft (11), fall into this category. For example, Clark found that his *Ss* recalled more of the affective passage (II) than they did of the neutral passage (III). However, the variances of these two passages differ significantly ($F = 1.84$, $P = .05$), with the affective passage being more variable. It is possible that greater variability in the number of words per IG enhances measured recall, thereby confounding Clark's results. The same consideration applies to the Levine-Murphy study in which both the pro-Communist and anti-Communist groups recalled the pro-Communist passage better than the anti-Communist passage. The variance ratio is again significant ($F = 2.27$, $P = .01$), with the pro-Communist passage having the greater variability.

In the light of the foregoing discussion, three experiments were undertaken. The purpose of the first was to determine whether a group of *Es*, guided by the same instructions, will divide passages in different ways. The purpose of the second was to investigate the effects of differences in mean words per IG on the mean recall score of a group of *Ss*, and on the relative position of *Ss* within the group. The aim of the third was to determine the effect of variability of the number of words per IG on the mean recall score and on relative position.

EXPERIMENT I

Method.—Two connected verbal passages of the Wechsler Memory Scale (12) and two passages from the Stone-Girdner-Albrecht alternate form of the scale (10) were presented to 31 faculty members and graduate students at the Iowa Child Welfare Research Station, with the following instructions: "Imagine that you are conducting an experiment in which it is necessary to obtain quantitative scores representing the recall of these four passages by a group of subjects. The purpose of your experiment is immaterial; all that matters is that you need an objective measure of the subjects' ability to recall the stimuli. Divide each passage into 'units' in some fashion. The procedure you use is entirely at your own discretion. There are no limitations or conditions except those already explicitly stated."

Only two members of the group had any prior acquaintance with the memory scales, and both later stated that their familiarity was insufficient to influence their performances.

Results.—The number of IG's per passage was counted for each *E*'s division of each passage. The data are shown in Table 2. Passages A and B are from the alternate form, C (reproduced on p. 0), and D are from Wechsler's original scale.

The significant variation among *Es* in dividing the passages, as well as among the passages themselves, is demonstrated by the analysis of variance of Table 3. Not only are there differences among *Es*, but the same *E* is likely to divide different passages so as to obtain different mean numbers of words per IG.

TABLE 3

ANALYSIS OF VARIANCE OF IDEA GROUPS OF FOUR PASSAGES BY 31 *Es*

Source	*df*	Mean Square	*F*	*P*
Es	30	128.77	23.58	.001
Passages	3	38.10	6.98	.01
$E \times P$	90	5.46		

Experiment II

Method.—To determine the effects of variations in mean number of words per IG on experimental results, Passage C was administered to a group of 29 *Ss* as a rote memory test. Each protocol was scored using four different divisions of the stimulus passage: Wechsler's own grouping (12) and three groupings selected from eighteen in Exp. I.[3] Grouping T had 11 IG's, which is apparently the smallest number which will yield a useable distribution of scores. Grouping U had 15 IG's, V had 24 IG's (the largest number made by any *E* in Exp. I), and W, Wechsler's division, also had 24.

Since the passage has 67 words, the groupings yield respective mean numbers of words per IG of 6.09, 4.47, 2.79, and 2.79. The corresponding *SD*'s are 1.97, 1.33, 1.41, and 1.38. An *F* ratio among means is 16.62, $P < .01$, while a Bartlett's test of the variances results in a χ^2 of 3.29, $P > .30$. Thus, the means differ significantly, but the variances are homogeneous.

Results.—The mean number of IG's recalled as a percentage of the number of IG's in the passage, and the *SD* of the distribution for the 29 *Ss* for each of the four scorings is shown in the left half of Table 4.

After converting the percentages in arcsins, an analysis of the variance yields an *F* of 219.27 for the means, and 18.32 for *Ss*. Both are significant beyond the .001 level. In general, the mean recall score tends to be inversely proportional to the mean number of words per IG, and directly proportional to the number of IG's. It is concluded that variations in the mean number of words per IG, apart from differences in variances, can result in significant differences in the recall score of a group of *Ss*.

The rank-order correlations among ranked recall scores for the four divisions are given in the right half of Table 4. The average rho is. 887. It is evaluated in the null case by means of a transformation[4] into Kendall's coefficient of concordance, *W* (4), which is in turn evaluated by means of χ^2_r. In this case, $W = .91$, and $\chi^2_r = 101.92$, which is significant beyond the .001 level for 28 *df*.

Evidently, differences in mean number of words per IG have very little effect on the relative positions of *Ss* within the group, despite the differences in mean recall scores.

TABLE 4

Mean Recall Scores in Percentages for Four Divisions of Passage C, and Rank Intercorrelations Among Them

Division	% Recall		Correlations		
	Mean	*SD*	U	V	W
T	22.73	16.45	.83	.93	.88
U	31.53	15.73		.88	.89
V	41.13	18.21			.91
W	41.96	17.29			

Experiment III

Method.—To investigate the effects of variability of the number of words per IG, four divisions of Passage C having identical means but heterogeneous variances were devised. These scorings, which are designed as X, Y, and Z, together with W, Wechsler's division, all contained 24 IG's, and hence each had a mean number of words per IG of 2.79. The respective variances were 4.61, 3.22, .61, and 1.91. A χ^2 obtained from a Bartlett's test was 16.61, significant beyond the .001 level. Thus the variances differ significantly, though the means are the same.

Passage C was again administered to a group of 25 *Ss* as a rote memory test.

Results.—Since the number of IG's is identical for each scoring, the mean

[3] Thirteen of the 31 divisions had less than 11 IG's. These divisions failed to yield useable distributions of scores. The highest obtained score seldom exceeded 3 IG's, and 50% or more of the scores were zero. It appears that when the division system has an average of 7 or more words per IG, the task of accurate recall of complete IG's becomes too difficult, and distributions of scores are flattened beyond utility.

[4] *W* is a linear function of the average rho, and can be obtained by the formula $W = \frac{\rho m - \rho + 1}{m}$ in which *m* is the number of rankings being averaged. It can also be computed directly from the rankings (cf. 4).

TABLE 5

MEAN RECALL SCORES IN IG'S FOR FOUR DIVISIONS OF PASSAGE C, AND RANK CORRELATIONS AMONG THEM

Division	IG's		Correlations		
	Mean	*SD*	X	Y	Z
W	10.92	2.99	.83	.90	.88
X	9.96	3.28		.79	.86
Y	10.60	3.05			.88
Z	9.88	2.92			

number of IG's recalled by the *Ss*, rather than a percentage, and the respective *SD*'s are listed in the left half of Table 5.

An analysis of variance yields a significant F of 4.93, $P < .005$, among means. The variance due to *Ss* is also significant, as expected ($F = 27.40$, $P < .001$). Differences in the variability of the number of words per IG can thus also produce significant differences in mean recall scores, although there is no clear relationship between size of the variance and the recall score.

The rank intercorrelations among the scorings are shown in the right half of Table 5. The coefficients are uniformly high, similar to those in Table 4. The average rho is .857 and W is .89, which is again highly significant. The conclusion is the same as that for Exp. II: differences in variability of the number of words per IG affects the relative positions of *Ss* only slightly, in spite of the differences in mean recall scores.

DISCUSSION

Three primary conclusions can be drawn from the data presented here: (*a*) different *Es* are likely to prepare connected verbal stimuli for quantitative experiments in different ways, and in addition, an individual *E* may himself divide passages differentially; (*b*) differences in the mean number of words per IG can result in significantly different mean recall scores for a sample of *Ss*; and (*c*) differences in variability of the number of words per IG can also produce different mean recall scores. As a secondary conclusion, we can say that neither type of difference in division is apt to affect relative positions of *Ss* in the group.

The implications of the experiments reported here were suggested in the introduction. In a study in which two or more passages are used as stimuli, or when experimental results are to be compared with those of an earlier work, the method of division of the stimuli may be a pertinent factor. Differences in division are likely to cause experimental findings to fluctuate, and will thus confound the effects of the experimental variable. As previously noted, many of the studies of mnemonic dynamics may have been subject to this type of confounding.

The results of Exp. II indicate a tendency for the mean recall score to vary inversely with the number of words per IG. Analysis of the data of Table 1 had suggested that greater variability of the number of words per IG makes for a higher recall score, but this conclusion was not substantiated by the results of Exp. III. However, differences in variability were uncomplicated by differences in mean number of words per IG in Exp. III. This was not true of the Levine-Murphy (**5**) and Clark (**2**) studies. This raises the possibility of a cumulative effect of differences in both means and variability; it is possible that effects are additive in some unknown fashion.

The "methods" for preparing connected verbal stimuli for quantification which have been reported in earlier studies usually do not actually stipulate rules for making the division into units. The experiments reported here point up the potential value of some real method for dividing stimuli so that the mean and variance of the number of words per IG do not vary from passage to passage. A

simple technique which appears to be suitable has been devised (6). It consists of dividing the passage into its grammatical entities. The subject of a sentence, the predicate, and prepositional and other clauses, each constitute an IG. The following sentence is divided into "grammatical idea groups": He/ went/ for days/ with very little sleep/ and is/ so tired/ that he/ can barely move/ a muscle/.

This method usually results in a mean of about 2.2 words per IG with an *SD* of about 1.00 for ordinary prose. A fine unit is probably the more efficient measure, but should it be deemed necessary the mean and variability can be increased somewhat by combining *to be* verbs with the predicate, and by adding pronouns to the subsequent verb, thus: He went/ for days/ with very little sleep/ and is so tired/ that he can barely move/ a muscle/.

If the stimulus passages are similar in reading ease as measured by any conventional formula, the use of the grammatical IG should result in divisions which do not differ significantly. Even numerical differences will usually be quite small.

Summary

It was determined by an analysis of previously published data, and by an experiment reported here, that different *Es* are likely to prepare connected verbal material for quantitative scoring in different ways. Experiments were performed to assess the effects of differences in division on experimental data. The results indicate that differences in either mean number of words per idea group, or in the variability of the number of words per idea group, can affect experimental findings significantly, and may obscure measurement of the experimental variable. Relative positions of *Ss* in a group are not affected by either type of difference. The use of the grammatical idea group was recommended as a method of preparing stimuli which will eliminate confounding differences in division.

References

1. Alper, T. S., & Korchin, S. J. Memory for socially relevant material. *J. abnorm. soc. Psychol.*, 1952, **47**, 25–37.
2. Clark, K. B. Some factors influencing the remembering of prose material. *Arch. Psychol., N. Y.*, 1940, No. 253.
3. Henderson, E. M. A study of memory for connected trains of thought. *Psychol. Monogr.*, 1903, **5**, No. 6 (Whole No. 23).
4. Kendall, M. G. *Rank correlation methods.* London: Charles Griffin, 1948.
5. Levine, J. M., & Murphy, G. The learning and forgetting of controversial material. *J. abnorm. soc. Psychol.*, 1941, **38**, 507–517.
6. Levitt, E. E. A quantitative investigation of individual differences in serial reproduction: a contribution to the study of rumor. Unpublished doctor's dissertation, Columbia Univ., 1952.
7. Lindquist, E. F. *Design and analysis of experiments.* New York: Houghton Mifflin, 1953.
8. Lyon, D. O. *Memory and the learning process.* Baltimore: Warwick & York, 1917.
9. Menzies, R. The comparative memory values of pleasant, unpleasant and indifferent experiences. *J. exp. Psychol.*, 1935, **18**, 267–279.
10. Stone, C. P., Girdner, J., & Albrecht, R. An alternate form of the Wechsler Memory Scale. *J. Psychol.*, 1946, **22**, 199–206.
11. Taft, R. Selective recall and memory distortion of favorable and unfavorable material. *J. abnorm. soc. Psychol.*, 1954, **49**, 23–28.
12. Wechsler, D. A standardized memory scale for clinical use. *J. Psychol.*, 1945, **19**, 87–95.

(Received July 12, 1955)

Journal of Experimental Psychology
Vol. 52, No. 1, 1956

EFFECT OF OVERLEARNING OF A VERBAL RESPONSE ON TRANSFER OF TRAINING[1]

GEORGE MANDLER AND SHIRLEY H. HEINEMANN

Harvard University

A previous study by Mandler (**5**) has described transfer effects as a function of prior overlearning of motor responses. The present experiment represents a replication of that study in a different, i.e., verbal, effector modality. In the preceding study *S*s were given training on a task which required the assembly of discrete motor units into a complex response. Degree of overlearning, i.e., practice on the training task beyond the stable establishment of S-R connections or habits, was related to subsequent transfer performance. The results indicated increasing positive transfer when a practiced response had to be associated to a new stimulus, initial negative transfer when a new response had to be learned to a previously used stimulus, and positive transfer at high degrees of overlearning when both stimulus and response had been used in the training task but not previously paired. While the first and second findings were consistent with previous work (**9, 11**), the last result conflicts with a theoretical analysis by Gagné, Baker, and Foster (**4**) and a confirmation of their prediction, with verbal material, by Porter and Duncan (**8**). These investigators predicted and found maximum negative transfer in the S-R reversal condition, i.e., when previously used stimuli and responses were paired in new combinations.

The present experiment is concerned with the generality of the previous findings for motor behavior. Here we have employed a task in which verbal units are assembled into complex overlearned responses. The study of three transfer situations is patterned after Bruce's paradigm (**2**). He also used verbal material, but restricted himself to lower degrees of training. This study should also provide a further test of the theoretical analysis of response factors in human learning (**6**).

METHOD

Task and apparatus.—The responses used were three-place consonant nonsense syllables selected from Witmer's standardization study (**13**). The stimuli were single integer numbers. The material was presented on a Hull-type memory drum. All *S*s, except one control group, were trained on an original task (Task I) and tested on a transfer task (Task II).

In both tasks the stimulus item for each pair was presented for 2.5 sec., followed by a 1.5-sec. exposure of stimulus and response together. Then the next stimulus appeared with the response window closed, and so forth. The *S* was required to pronounce the three letters, and a response was scored correct if it was elicited by its stimulus at any time prior to the exposure of the response on the memory drum. A complete record of all responses was kept by *E*.

Subjects.—The *S*s were 60 paid undergraduate students obtained through the student employment office. They were assigned at random to 10 groups (five degrees of training and two forms) as they appeared for the experiment.

Instructions.—All *S*s were given standard paired-associate instructions prior to Task I. They were also told that in the course of the experiment *E* might stop the procedure in order to make some adjustments to the apparatus.

Transfer conditions.—The *S*s learned a four-pair training task (Task I) to the various degrees of overlearning and then were switched to an eight-pair transfer task (Task II), which simultaneously tested transfer effects for four con-

[1] This study was supported by a grant from the Laboratory of Social Relations, Harvard University.

TABLE 1

STIMULI AND RESPONSES USED IN THE TWO FORMS OF TASKS I AND II

Transfer Conditions Represented in Task II	Form A				Form B			
	Task I		Task II		Task I		Task II	
	S	R	S	R	S	R	S	R
I. (Old S—New R)	7	LFH (50)*	7	XSK (63)	2	GBL (54)	2	MJS (54)
	9	KGR (71)	9	HMQ (54)	5	FRJ (58)	5	BQX (58)
III. (Old S—Old R)	3	BQX (58)	3	LFH (50)	8	XSK (63)	8	GBL (54)
	4	MJS (54)	4	BQX (58)	6	HMQ (54)	6	XSK (63)
II. (New S—Old R)			5	MJS (54)			7	FRJ (58)
			2	KGR (71)			4	HMQ (54)
IV. (New S—New R)			6	GBL (54)			9	LFH (50)
			8	FRJ (58)			3	KGR (71)

* Figures in parentheses show the association values obtained from Witmer (13). The mean association values of all the syllables used in each condition are: Cond. I: 57.75, Cond. II: 58.75, Cond. III: 56.75, Cond. IV: 58.25.

ditions of transfer. The stimuli and responses used in Task I will be called "old" stimuli and responses; those first introduced in Task II will be called "new." This results in four transfer conditions:

I. Learning to make a new response to an old stimulus.

II. Learning to make an old response to a new stimulus.

III. Learning to make an old response to an old stimulus when these have not been previously paired. This involves the recombination of stimuli and responses used in Task I.

IV. Learning to make a new response to a new stimulus. This condition was used as a control for such phenomena as the warm-up effect demonstrated by Thune (10).

Two forms (A and B) of Tasks I and II were used. Table 1 shows the stimuli and responses used in both forms of the two tasks. Twelve different letters were selected in such a way as to provide maximum phonemic differentiation. The responses were obtained from Witmer's lists (13). They fell within as narrow a middle range of association values as was feasible within the other requirements of the design. The mean association values (shown in Table 1) of the responses used in the various conditions did not differ significantly from each other. The letter combinations also fulfilled the following conditions: (*a*) Any subsequence of two letters that was used in a given response was not used in any other response. (*b*) In Task I, no letter was used more than once; in Task II, no letter was used more than twice.[2] (*c*) No letter appeared more than once as either the first, second, or third member of a set. (*d*) In those transfer conditions (I and III) where two interfering responses are elicited by the same stimulus, no letter is common to both responses.

A "trial" in Task I consisted of a repetition of the four S-R pairs. Twelve different orderings of pairs within trials were used, the only restriction on the ordering being that the first pair of a succeeding trial be different from the last pair of the preceding trial. This gave a total sequence of 48 pairs (12 trials) which was successively repeated. A comparable method was used with the transfer task except that with eight pairs per trial, six trials completed a run-through.

Degree of training conditions.—The degrees of training used in this experiment were the same as those used in the motor study. They consisted of 0, 10, 30, 50, and 100 errorless trials on Task I. A different group of *S*s was used for each of these conditions; these five groups will hereafter be designated in terms of the amount of training received.

Group 0 had no training on Task I. In Group 10, *S* was stopped on Task I as soon as

[2] A preliminary investigation showed that it would not be possible to parallel the motor study more exactly by using combinations of only six different letters. None of five pilot *S*s was able to learn a list of eight stimulus response pairs where the eight different responses were made up of only six different letters.

his performance included 10 errorless trials. In Group 50, there was a rest for 1.5 min. after 25 errorless trials, and then *S* continued until he reached the criterion of 50 errorless trials. In Group 100, *S* was treated the same as were Group 50 *Ss*, except that when he finished 50 errorless trials he was told to come back the next day. On the second day, he was given two additional blocks of 25 trials each on Task I. At the conclusion of the requisite number of errorless trials, all *Ss* were given a 3-min. rest, while *E* changed the tape on the memory drum. They then worked on Task II to a criterion of two successive errorless trials (16 correct responses) or 20 trials, whichever occurred last. Group 0 proceeded, after identical initial instructions, directly to Task II.

Results

Training results.—An analysis of variance of the training data of all *Ss* who were given Task I, using number of trials to the first correct trial, showed no significant differences between groups. There was a significant difference between Forms A and B. Form A had a mean of 17.5 (SD = 12.5) and Form B a mean of 10.5 (SD = 6.6) trials to the first correct trial.

Figure 1 shows the mean number of errors in successive blocks of trials containing 10 correct trials for both forms, which were combined since no significant differences were found between forms from 20 correct trials on. In Fig. 1 all four training groups were combined for the first point; Groups 30, 50, and 100 are combined for the second and third point, and so on. It can be seen that the number of errors reaches asymptotic values by 40 correct trials. It should be pointed out that the mean error values are extremely low. For the 50 correct trial point the average *S* made .33 errors, while giving 40 correct responses (four pairs each of 10 correct trials). From that point on the error percentage is less than 1%, with the exception of the block of 10 correct trials immediately following the 24-hr. rest. Thus, at least for Groups 50 and 100, overlearning, defined as training beyond the establishment of stable S-R connections, can be presumed to have been established.

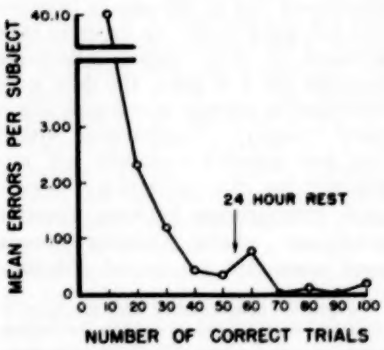

Fig. 1. Mean errors per subject to perform successive blocks of ten perfect trials, shown as a function of level of training (number of correct trials achieved).

Transfer results.—Two different transfer measures, identical to the ones used in the previous study, were obtained for each *S* for each of the three transfer conditions. Condition IV was used as a control. One measure, the F-20 score, used the frequency of correct responses during the first 20 trials on Task II. The raw frequency on the control condition was subtracted from the frequency on each of the relevant transfer conditions. The other measure, the FCR score, was the logarithm of the ratio of transfer to control performance in terms of the number of the trial on which the first correct response occurred. These measures use *Ss* as their own controls and utilize control conditions that are specific to the particular form used.

Analyses of variance of the transfer measures showed no significant difference associated with the interaction between Forms and Groups, or be-

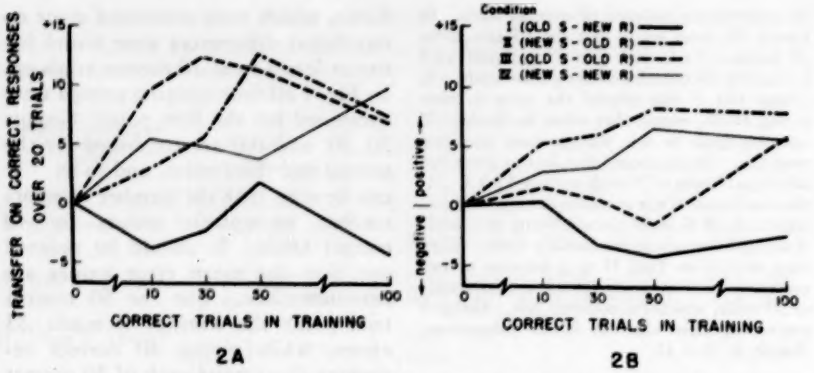

FIG. 2. Transfer of training as a function of degree of original training in terms of frequency of correct responses in 20 transfer trials (F-20). Fig. 2A shows results of the present study, Fig. 2B those of the motor study.

tween Forms, Groups, and Conditions; the data were therefore combined. The analyses did show a significant effect due to Conditions ($P < .001$), as well as significant interactions between Groups and Conditions ($P < .001$). The P values presented below refer to the significance of t tests based on mean scores.

Figures 2A and 3 show the results for the two transfer measures, F-20 and FCR. Figure 2B shows the results of the motor study for comparison purposes.[3] In Fig. 2A and 2B the results for the control group (Cond. IV) have also been included, but note that the curves for the other conditions represent mean *differences* between the particular score for that condition and the score on Condition IV. As in the motor study, Group 0, without previous training, showed some differences between the four conditions due to differences between the S-R pairs used. In order to take account of these initial differences between the S-R pairs, the data were corrected to assume a common origin (zero transfer). Comparisons within any one transfer condition are unaffected by this correction. Statistical comparisons between transfer conditions among training groups were controlled for initial difficulty

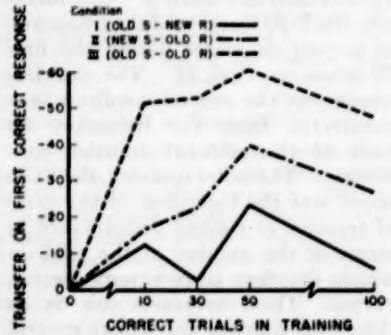

FIG. 3. Transfer of training as a function of degree of original training in terms of attainment of the first correct response (FCR). All transfer effects are positive; the negative values are a function of the measure used: $10 \log \frac{\text{Transfer score}}{\text{Control score}}$.

[3] Figure 2B differs slightly from the figure in the motor study due to an error in the original graph. Cond. I, Group 100, shows some negative transfer as in Fig. 2B, rather than zero transfer. This drafting error does not affect any of the results or discussion in this or the previous study.

by testing the significance of differences in all such comparisons against the differences obtained in the corresponding comparisons with Group 0.

Before discussing the specific transfer results, one finding should be pointed out. In both Fig. 2A and 3 the curves for all three transfer conditions show a parallel drop from Group 50 to Group 100. Since all the transfer data have been corrected for Cond. IV, it is obvious from Fig. 2A that the drop for the three transfer conditions can be accounted for completely by the rise on the Cond. IV (New S—New R) curve. This relationship also holds for the FCR measure. Thus, while no change took place for the three transfer conditions with an increase of 50 correct trials, the facilitative effect of these additional trials for the control condition makes for an apparent drop in the other conditions. No such simple explanation holds for any of the other findings, where the three transfer conditions do not show identical changes as a function of the control condition; nor is this explanation applicable to the motor response findings as shown in Fig. 2B. In the description of the differential transfer results we will therefore not be concerned with this change from Group 50 to Group 100, the magnitude and direction of which can be directly accounted for by the control effects.

Learning a new response to an old stimulus (Cond. I) shows different effects for the two measures. For F-20 there is an initial negative transfer effect followed by a return to about zero transfer. However, none of the differences is significant. For the FCR measure we find no negative transfer effect at all, but rather a significant positive transfer effect from 0 to 50 ($P < .01$) and from 30 to 50 ($P < .02$).

In Cond. II, learning an old response to a new stimulus, we find increasing positive transfer as a function of increasing training. This effect, however, does not reach statistically significant proportions until the 50 trial point ($P < .01$ for both measures). In this condition 21% of the responses in Group 100 were correctly given by the second trial, i.e., showed one trial learning.

Learning to make an old response to an old stimulus when these have not been previously paired (Cond. III) shows the highest positive transfer effects. Positive transfer is already significant for Group 10 ($P < .05$ for F-20, and $P < .01$ for FCR), and remains at that high level for the other groups. This condition also shows the highest degree of one trial learning, i.e., 29% of the responses in Group 100.

In comparing conditions within groups we find that on the F-20 measure (Fig. 2A) all three conditions differ significantly from each other in Group 10 at the .02 level or better. This difference is maintained until Group 50 where Cond. II and III are significantly better than Cond. I ($P < .001$), but do not differ from each other. On the FCR measure (Fig. 3) Cond. III is consistently superior to the other two conditions ($P < .05$ or better); Cond. I and II differ significantly from Group 30 on.

In comparing these results with those of the motor study, Fig. 2A and 2B reflect the major differences. Of primary interest is the superiority of Cond. III (Old S—Old R) to the other conditions in the present study. In the motor study Cond. III did not show any significant positive transfer until Group 100, and for the F-20 measure it was consistently inferior to Cond. II (Fig. 2B). The present experiment shows that Cond. III not

only shows positive transfer from Group 10 on, but also is superior to Cond. II in Groups 10 and 30, and essentially equal to it in Groups 50 and 100. Apart from this striking change in Cond. III there is some indication of a difference in Cond. I. In the present results we found substantially a lack of any significant transfer effect. In the motor study the F-20 measure also showed no significant effects, but for the FCR measure negative transfer approached statistical significance for Group 30. The present data show no negative transfer at all for this measure (Fig. 3).

The different results of the two studies can be summarized as a general trend toward positive transfer or increase in positive transfer for all three conditions in this study, and a strikingly superior positive transfer effect for Cond. III.

Warm-up results.—The warm-up effect refers to a general improvement in performance as a function of increasing experience in the experimental situation. Condition IV (New S—New R) is a direct measure of this effect and shows a consistent increase as indicated in Fig. 2A. A more detailed discussion of the warm-up effects in the motor study (**7**) shows that the Group 100 performance was inferior to the Group 50 performance. In the present study we find a reverse of this effect, i.e., an increase from Group 50 to Group 100. While this increase is not significant, the difference does account for the depressing effect on the three transfer conditions.

Discussion

Our results show gross similarity to the previous findings with motor responses (**5**). The increasing positive transfer for Cond. II and III was predicted on the basis of response integration during training. During training parts of complex responses serve as stimuli for each other and increasing correct repetitions during overlearning of these responses increase the integration and cohesiveness of the responses; the *S* learns how to perform the response. When these responses are then required in a new situation they are more easily elicited and maintained; *S* need not learn the response complex again. The lack of negative transfer in Cond. I (Old S—New R) has been ascribed to an intervening process. We have postulated that overlearning produces symbolic analogues of the overt response. The symbolic responses are most stable when the response is highly overlearned, and they make implicit trial-and-error behavior possible. This intervening process facilitates the inhibition and elimination of an incorrect response before it becomes overt.

The results of Cond. I can be compared with studies by Underwood (**11**) and Atwater (**1**). In both of these studies, which used the A-B, A-C paradigm, the results showed increasing facilitation as a function of degree of prior learning. In neither of our studies have we found any consistent positive transfer for this condition. Two differences should be pointed out. Both Underwood and Atwater used meaningful, previously practiced responses, namely adjectives and three-letter words, respectively. Neither of the studies controlled for warm-up effects which, as we have shown, consistently produce positive transfer. A control for this latter factor alone would probably significantly decrease their positive transfer findings.

Condition II is consistent with Wylie's (**12**) and Bruce's (**2**) data, and with the generalization that learning an old response to a new stimulus produces positive transfer.

In the discussion of the results of the motor study we pointed out (**5**, p. 416) that our findings for Cond. III were inconsistent with Gagné, Foster, and Baker's (**4**) analysis of the "stimulus-response reversal" condition, and Porter

and Duncan's (8) experimental evidence. These investigators had found maximum negative transfer for this condition. Our present data are even more in contradiction to that finding in that we found maximum *positive* transfer for this condition. In contrast with our study, Porter and Duncan used adjectives as stimuli and responses, pretrained their subjects to only one correct trial, and tested only one transfer condition within any one list. However, with familiar stimuli, i.e., adjectives, where response integration is already established, we would have predicted relatively little negative transfer. On the other hand, little or no response learning takes place when such familiar highly practiced responses are used, and other, interfering, processes may more easily come to the fore. It might also be possible that transfer effects associated with very high degrees of training are distinctively different from those associated with relatively lower degrees of learning. Finally, in Porter and Duncan's study all S-R pairs in any one list tested either the old stimulus-old response or the old stimulus-new response variation. In the present study only two out of eight pairs represented each condition. Thus, for the old stimulus-old response condition the *Ss*' previous experience with both the stimulus and the response elements might have facilitated recognition of this relationship, made discrimination of these two pairs easier, and accelerated their acquisition. Such an explanation of the divergence of the two findings would require an additional assumption, namely, that the recognition of an old stimulus-old response relationship is more advantageous for acquisition in a heterogeneous list, such as ours, than in a homogeneous list, such as Porter and Duncan's.

One final difference between the two overlearning studies should be noted. In the first place, transfer effects are more pronounced with verbal responses; secondly, additional training beyond 50 correct trials produced no differential transfer effect other than that accounted for by the increased warm-up facilitation. This finding is consistent with the generally accepted notion that verbal responses are the most highly overlearned human responses. Human *Ss*, particularly college students, have high degrees of experience in the novel recombination of letters and should show greater and more rapid transfer effects than with the assembly of motor responses. In addition, fewer trials are apparently necessary to bring such novel verbal combinations to an asymptotic level of integration. Our finding of significant positive transfer effects associated with high degrees of overlearning is consistent with a recent analysis of overlearning and ability by Ferguson (3). Since adult human behavior is primarily constituted of overlearned responses, and associated with symbolic activity, a further investigation of these processes should shed more light on individual differences in the performance of familiar tasks and responses.

Summary

Sixty *Ss* learned three-place consonant combinations in response to single number integers in both a training and transfer task. The degrees of training were 0, 10, 30, 50, and 100 errorless trials on the training task, with 12 *Ss* in each of these groups. The transfer task simultaneously tested four different conditions, three transfer and a fourth condition which controlled for warm-up. The results were:

1. Learning to make a new reponse to an old stimulus showed no consistent negative or positive transfer effects.

2. Learning to make an old response to a new stimulus showed increasing positive transfer as degree of original training was increased.

3. Learning to make an old response to an old stimulus, when these had not been previously paired, also showed increasing positive transfer as original training was increased, but a significantly higher degree of positive transfer than the New S—Old R condition.

These data were discussed in relation to a previous comparable study which used motor responses.

References

1. Atwater, S. K. Proactive inhibition and associative facilitation as affected by degree of prior learning. *J. exp. Psychol.*, 1953, **46**, 400–404.

2. Bruce, R. W. Conditions of transfer of training. *J. exp. Psychol.*, 1933, **16**, 343–361.
3. Ferguson, G. A. On learning and human ability. *Canad. J. Psychol.*, 1954, **8**, 95–112.
4. Gagné, R. M., Baker, K. E., & Foster, H. On the relation between similarity and transfer of training in the learning of discriminative motor units. *Psychol. Rev.*, 1950, **57**, 67–79.
5. Mandler, G. Transfer of training as a function of degree of response overlearning. *J. exp. Psychol.*, 1954, **47**, 411–417.
6. Mandler, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.
7. Mandler, G. The warm-up effect: some further evidence on temporal and task factors. *J. gen. Psychol.*, in press.
8. Porter, L. W., & Duncan, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.
9. Siipola, E. M., & Israel, H. E. Habit interference as dependent upon stage of training. *Amer. J. Psychol.*, 1933, **45**, 205–227.
10. Thune, L. E. The effect of different types of preliminary activities on subsequent learning of paired-associate material. *J. exp. Psychol.*, 1950, **40**, 423–438.
11. Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34.
12. Wylie, H. H. Transfer of response in the white rat. *Behav. Monogr.*, 1909, **3**, No. 5.
13. Witmer, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–359.

(Received July 26, 1955)

Journal of Experimental Psychology
Vol. 52, No. 1, 1956

SIMULTANEOUS AND SUCCESSIVE DISCRIMINATION IN A SINGLE-UNIT HOLLOW-SQUARE MAZE

ALLEN D. CALVIN[1] AND CLARENCE M. WILLIAMS

Michigan State University

In a recent experiment Weise and Bitterman (**6**) compared the relative difficulty of simultaneous and successive discrimination using a multiple discrimination apparatus consisting of four hollow-square units. They reported that the successive discrimination was easier than the simultaneous. Some experimental work with children by Calvin (**2**) indicated the possibility that the anomalous results obtained by Weise and Bitterman were due to the fact that the stimuli were presented in such close spatial contiguity that they appeared as a single unified percept rather than as discrete units. Bitterman and Wodinsky (**1**) have presented a similar explanation of the Weise-Bitterman findings. However, when this hypothesis was tested by Calvin and Seibel (**3**), it was not confirmed. The simultaneous discrimination remained relatively more difficult than the successive even when the stimuli were spatially discrete. Calvin and Seibel further reported that various factors such as goal gradient, alternation, etc. were operating differentially in the two types of discrimination. The present experiment was designed to investigate further the relative difficulty of simultaneous and successive discrimination using only one unit of the Calvin-Seibel modified multiple discrimination apparatus.

METHOD

Apparatus.—Our apparatus duplicated Calvin and Seibel's (**3**) with the exception that only one unit of the four-unit hollow-square maze was used. As in their experiment, two lamps served as discriminanda. Each was mounted 8 in. from the choice point. The starting box was 8 by 8 in., the food box 8 by 17 in., and the alleys were 12 in. long and 4½ in. high. The apparatus was painted a flat black and the top was covered with hardware cloth. The experimental room was dimly illuminated by an infrared bulb.

Subjects.—The *S*s were 20 naive albino female rats approximately 3 mo. old from the colony maintained by the psychology department at Michigan State University. Two *S*s were discarded on the first day because of an error by *E*, and two spare *S*s which had also been handled along with the main group were substituted. During the preliminary gentling it was discovered that a male *S* had inadvertently been placed in one of the experimental cages. Although he was removed immediately, it later became apparent that two of the *S*s had been impregnated. They were discarded and two more *S*s were substituted for the impregnated *S*s.

Procedure.—The *S*s were randomly divided into two groups. Group I learned the simultaneous problem and Group II the successive. For the simultaneous group either the left or right bulb at the choice point was lighted. Half the *S*s in this group had light positive and half dark positive. The position of the positive cue was varied according to a predetermined random order with the restriction that the same side was never correct more than twice in succession and each side was correct for half the trials on each day. The *S*s were given eight trials a day and the criterion was two errorless days in succession.

For the successive group the bulbs at the choice point were either both lighted or both dark. Half the *S*s had to go left to light and right to dark and the other half had to go left to dark and right to light. The same random orders that were used for the simultaneous group were employed, and again the criterion was two errorless days in succession.

During the first few days of training some *S*s occasionally refused to move, and when this occurred *E* prodded them. In no case was prodding necessary beyond Day 3. As in the Calvin-Seibel and Weise-Bitterman procedure, retracing was allowed. For a more detailed

[1] Now at Hollins College.

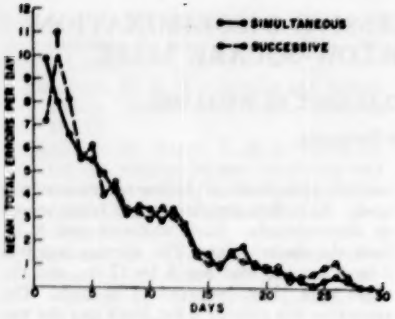

FIG. 1. Mean total error scores.

description of the procedure see Calvin and Seibel (3).

RESULTS

Since retracing was allowed, *S* could make more than one error on a single trial. Figure 1 shows the course of learning in terms of total errors. (The curves for initial errors were, as would be expected, quite similar.) As may be seen, there is no marked difference between the two groups. When they are compared by means of the Mann-Whitney *U* test (4), the difference between them falls far short of significance for both total and initial errors.

In Fig. 2 the results for Group I, the simultaneous group, are broken down in terms of those *S*s who had light positive and those *S*s with dark positive and these two groups are

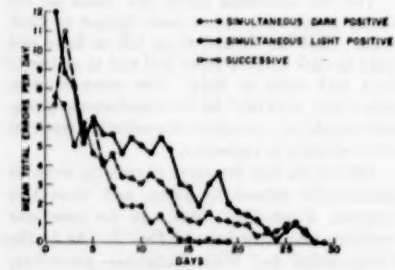

FIG. 2. Mean total error scores.

TABLE 1

MEAN ERROR SCORES

Group	Initial Errors	Total Errors
Simultaneous light positive	92.2	104.8
Simultaneous dark positive	36.8	46.6
Successive	63.4	77.6

compared with the performance of the *S*s in Group II in terms of total errors. (A breakdown in terms of initial errors gave similar curves.) Table 1 presents mean error scores for the successive group and the two simultaneous subgroups. It appears quite obvious that the dark-positive simultaneous *S*s are markedly superior to the successive *S*s while the light-positive simultaneous *S*s are considerably inferior in performance when compared with the successive. The following tests of significance were made with the Mann-Whitney *U* test (one-tailed): the difference between the dark-positive simultaneous and the light-positive simultaneous is significant beyond the 1% level for both total and initial errors; the difference between the dark-positive simultaneous and successive is also significant beyond the 1% level for both total and initial errors; and the difference between the light-positive simultaneous and the successive is significant beyond the 5% level for both total and initial errors.

DISCUSSION

The fact that there was no significant difference in performance between Groups I and II is in direct opposition to the findings of Calvin and Seibel (3) and Weise and Bitterman (6), who found the successive discrimination to be relatively easier than the simultaneous. We used only a single unit while Weise and Bitterman and Calvin

and Seibel used a multiple-discrimination apparatus. The difference between our results and theirs supports Spence's contention (**5**), and the Calvin and Seibel findings, that using the multiple-discrimination apparatus makes an adequate analysis of simultaneous vs. successive discrimination extremely difficult. However, even with the confounding effects of the multiple-discrimination apparatus removed, the simultaneous *S*s were still not superior to the successive as would be expected from the approach-avoidance position.

A solution to the problem of the failure of *S*s of Group II to surpass those of Group I is suggested by an experiment conducted by Davis.[2] Using a delay situation in a multiple-discrimination apparatus with slightly longer arms, Davis found that his dark-positive simultaneous *S*s' performance was superior to that of his successives and his light-positive *S*s were inferior (although not significantly so) to his successives. A similar analysis for our data substantiated Davis's findings. Weise and Bitterman also found that the simultaneous dark-positive *S*s were superior to the light-positive although the difference failed to reach significance. However, even the performance of their simultaneous dark-positive *S*s taken by itself was significantly inferior to that of their successive *S*s. An analysis of Calvin and Seibel's data also reveals a tendency for the simultaneous dark-positive to be superior to the light-positive, but as in the case of Weise and Bitterman the difference was not significant and the simultaneous dark-positive *S*s were significantly inferior when compared to the successive. Thus the negative phototropic effect seems to be less important in the standard multiple-discrimination apparatus problem than in our single unit or Davis's delay condition.

If the phototropic effects were exactly balanced, i.e., if *S*s who had to avoid light (dark positive) were aided exactly as much as *S*s who had to approach light (light positive) were handicapped, then the phototropic effect would cancel itself out. However, as Davis suggests, the assumption of a balanced effect seems hardly tenable in the light of Spence's finding (**5**, p. 91) that when confounding factors such as phototropic effect, multiple discrimination, etc. were removed, the simultaneous discrimination was easier than the successive. In view of Spence's results and his further comment that "The finding of Weise and Bitterman was certainly a surprise to the present writer, for in the course of the last 15 years a number of studies had been run in the Iowa laboratories that had tended to suggest that the *successive* problem was relatively the more difficult of the two situations," it would seem highly probable that the simultaneous light-positive *S*s are handicapped more than the simultaneous dark-positive *S*s are aided, which would account for the lack of superiority of the simultaneous group as a whole when they are compared with the successive.

Since the only demonstrations of the relative ease of successive discrimination when compared to simultaneous have involved extremely complex apparatus (the four-unit multiple-discrimination apparatus and the multiwindow jumping stand) with their concomitant confounding of the analysis, it would seem that such demonstrations cannot be considered as crucial proof of the inadequacy of the approach-avoidance formulation.

[2] R. H. Davis. Personal communication, April 7, 1955.

Summary

Twenty albino rats were run in a single-unit hollow-square maze where lights were used as cues as in the Weise-Bitterman and Calvin-Seibel experiments. Half were presented with a simultaneous discrimination and half with a successive. There was no significant difference in terms of errors between the two groups. A further breakdown of the simultaneous group into those who had to approach light and those who had to avoid light revealed that those who had to approach light were significantly inferior to the successive animals and those who had to avoid light were significantly superior to the successive. Implications of these findings were discussed.

References

1. Bitterman, M. E., & Wodinsky, J. Simultaneous and successive discrimination. *Psychol. Rev.*, 1953, **60**, 371–376.
2. Calvin, A. D. Configurational learning in children. *J. educ. Psychol.*, 1955, **46**, 117–120.
3. Calvin, A. D., & Seibel, J. L. A further investigation of response selection in simultaneous and successive discrimination. *J. exp. Psychol.*, 1954, **48**, 339–342.
4. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
5. Spence, K. W. The nature of response in discrimination learning. *Psychol., Rev.*, 1952, **59**, 89–93.
6. Weise,P., & Bitterman, M. E. Response selection in discriminative learning. *Psychol. Rev.*, 1951, **58**, 185–195.

(Received July 12, 1955)

Journal of Experimental Psychology
Vol. 52, No. 1, 1956

STIMULUS UNITS AND RANGE OF EXPERIENCED STIMULI AS DETERMINANTS OF GENERALIZATION–DISCRIMINATION GRADIENTS [1]

JACOB L. GEWIRTZ, LYLE V. JONES, AND KARL-ERIK WAERNERYD

The University of Chicago

The slope and symmetry of the typical stimulus generalization gradient play important roles in a number of behavior theories (**4, 5, 9, 12**). Characteristics of the gradient generally are introduced as postulates, without direct experimental verification.

The form of a stimulus generalization gradient is a function of many conditions, some of which have been recognized in the literature (cf. **2**, p. 319; **11**), but few of which have been subjected to formal investigation. The purpose of the present study is to investigate, in human adults, the effect of (*a*) psychological vs. physical units of measurement of the stimulus scale, and of (*b*) central tendency, determined by earlier experience with the stimulus dimension, upon slope and symmetry of empirical gradients of response strength.

By stimulus generalization is meant the process whereby stimuli along a dimension of similarity with a training stimulus, and not associated in prior training with its learned response, tend to elicit the response. In contrast, by stimulus discrimination is meant the process whereby, as a function of differential reinforcement, stimuli other than the training stimulus tend not to elicit the learned response. There can be postulated a dimension of experimental conditions, ranging from those which yield pure stimulus generalization to conditions which yield pure stimulus discrimination (**1**). Gradients of response which fall more steeply from a maximum at the training stimulus are expected as the training conditions approach those of discrimination. The effects investigated in the present study are expected to appear in response curves at all points along the postulated generalization-discrimination dimension, excepting the extremes, where curves of minimal or maximal slope are expected. Experimental conditions of the present study represent an intermediate position along this postulated dimension.

From the time of Fechner's work in psychophysics, a distinction has been recognized between psychological and physical stimulus dimensions. In its most primitive form, the relation between the two might be characterized by postulating that the psychological scale (R) is a negatively accelerated function of the physical scale (S). This postulate is consistent with typical findings concerning psychological and physical units, and avoids the dilemma posed by evidence that Fechner's Law ($R = k \log S$) does not hold for all values of S.

An effective test of the use of psychological vs. physical stimulus units upon both slope and symmetry of response curves requires comparison of differential responses to stimuli near zero in magnitude with differential responses to stimuli far from zero in magnitude. The stimulus selected, that of visual angle, exhibits the essential property of zero

[1] This study was supported by a grant from the Social Science Research Committee of the University of Chicago.

magnitude, and has the added advantage of allowing for stimuli both to the right and left of zero angle, equivalent in magnitude, differing only in sign.

By Central Tendency is meant the tendency for *S*'s judgment of a stimulus magnitude to err in the direction of a central magnitude of previously experienced stimuli. While in psychophysical studies central tendency phenomena frequently have been considered explicitly (cf. **2,** pp. 323–335; **3, 6**), it is the rare analysis of generalization or discrimination data which has taken them into account **(e.g., 10)**.

Method

The voluntary response of "same" or "different" and a stimulus dimension of visual angle are employed. Each of four experimental groups is presented with two tasks: (*a*) training to respond in the center stimulus range to the zero angle stimulus, followed by testing for response to the training stimulus and to 10 stimuli symmetrically positioned around zero at 1-mm. intervals,[1] and (*b*) training to respond to a stimulus markedly different from zero angle, either a positive angle in the right range (+ 50 mm.) or a negative angle in the left range (− 50 mm.), followed by testing for response to that training stimulus and to 10 stimuli symmetrically positioned at 3-mm. intervals around it. Stimulus units for center and off-center ranges were selected on the basis of a preliminary study, results of which confirmed the expectation under the postulated relation between psychological and physical stimulus dimensions: to insure obtaining sufficient evidence concerning forms of the response gradients, it is necessary to employ a wider segment of the stimulus dimension in the off-center range than in the center range.

The four experimental groups differ both according to the order of presentation of the tasks, and the position of the off-center training stimulus (positive or negative angle, see Table 1). The postulated relation between psychological and physical units is expected to yield (*a*) steeper slope of response gradients at the center than at the off-center stimulus ranges, and (*b*) skewness of curves within off-center ranges in the direction away from zero visual angle. Central tendency is expected to affect response gradients from the second task, yielding curves displaced in the direction of the stimulus range in the initial task.

Apparatus.—The stimulus object was a ½-in. high movable pointer needle, illuminated by a prefocused 2.5-v. bulb. The needle was seen by *S* through a horizontal slot ¼ in. by 35 in., which was vertically centered at eye level on a plywood panel. The pointer needle was visible to *S* through the eyepiece of a right-eye viewing tube, 82 in. from the stimulus panel. A shutter on the viewing tube was operated by a solenoid.

Mounted between and directly in line with the viewing tube and the base of the pointer needle at its most central position were two sights, for definition of the zero angle stimulus. Each was a luminous sphere, ¾ in. in diameter, illuminated by a 2.5-v. prefocused bulb. At *S*'s position were two response keys, one for each hand, resting at an inclining angle upon microswitches.

Procedure.—With hands on the two response keys, *S*, seated in a position which aligned the fixed sights with the zero angle stimulus, was informed that the "strip of light" might appear in any of a number of positions. With shutter open, the stimulus was presented in eight positions: − 80, − 60, − 40, − 20, + 20, + 40, + 60, and + 80 mm. The *S* was informed that his "speed of judgment" was to be tested, and that his task was to release either the right or left response key as soon as the shutter was opened, depending upon whether the stimulus light was in a position "same as" or "different from" the "standard" (training stimulus). The response task was reversed for half the *S*s in each group. (For every analysis in this paper, the difference between right- and left-hand response keys as indicators of "same" responses has been tested. Because in no case is statistical significance approached by an effect involving response hand, that variable is subsequently ignored.) The *S*'s response served to close the shutter, which then remained closed between trials. Speed of response was emphasized.

After five 5-sec. presentations of the standard training stimulus, during which *S* was not to respond, there were 12 training trials, 10 demanding response to the training stimulus and 2 (No. 6 and 8) demanding response to stimuli 30 mm. to its right (Training 1 in Table 1). Without interruption, 55 test trials (Test 1)

[1] Stimulus intervals are given in terms of distance, in millimeters, of the pointer needle from its central, zero angle position. Throughout the stimulus range, 1 mm. transcends a stimulus angle of 0° 1′ 36″. The extremes of stimuli employed for any *S* transcend a visual angle falling within the range of foveal vision.

TABLE 1

DETAILS OF EXPERIMENTAL DESIGN

Group	Step			
	1 Training 1	2 Test 1	3 Training 2	4 Test 2
CR	10 of 12 trials to zero angle	55 trials to center range	10 of 12 trials to + 50-mm. angle (to right)	55 trials to right range
CL	10 of 12 trials to zero angle	55 trials to center range	10 of 12 trials to − 50-mm. angle (to left)	55 trials to left range
RC	10 of 12 trials to + 50-mm. angle (to right)	55 trials to right range	10 of 12 trials to zero angle	55 trials to center range
LC	10 of 12 trials to − 50-mm. angle (to left)	55 trials to left range	10 of 12 trials to zero angle	55 trials to center range

followed. Order of stimuli was randomized with the restriction that every one of the 11 stimulus positions within the appropriate stimulus range be represented once in each of the five cycles of trials.

A 5-min. rest period followed Test 1, after which the final procedures were instituted (Training 2 and Test 2). After *S* was seated and had aligned the sights with the zero angle stimulus, he was informed that the "speed of judgment" test was to continue, and that he would now see the "standard" (training stimulus) in a position different from those used before. The Training 2 and Test 2 procedures which then followed were the same as those employed for the Training 1 and Test 1 procedures, except that a different stimulus range was involved. The sequences of the training and test conditions for the four experimental groups are detailed in Table 1.

Subjects.—The *S*s were 21 men and 19 women, summer-session students at the University of

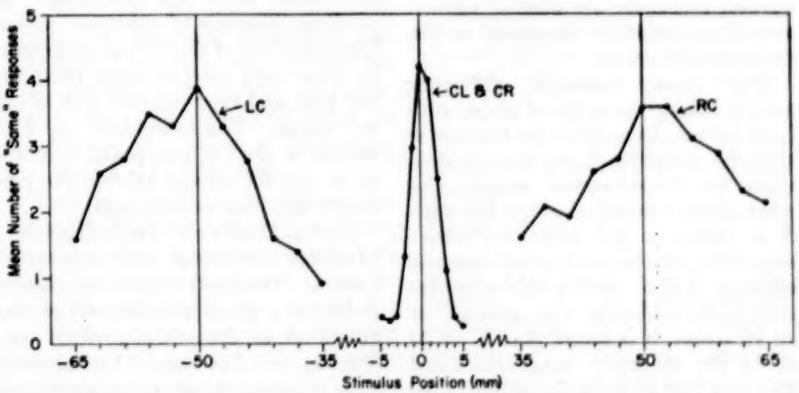

FIG. 1. Mean response curves over the three ranges of stimuli, Test 1.

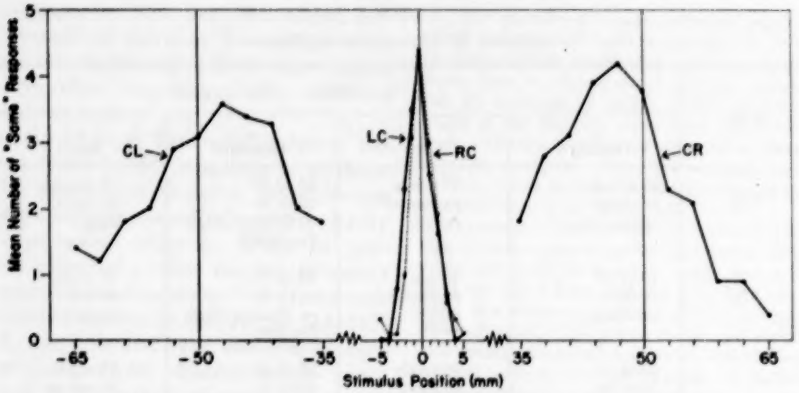

FIG. 2. Mean response curves over the three ranges of stimuli, Test 2.

Chicago, who were randomly assigned to the four experimental groups of 10 *Ss* each.

RESULTS AND DISCUSSION

Psychological vs. physical stimulus dimensions.—In Fig. 1 appear three mean response curves for the initial test (Step 2, Table 1) of the LC, RC, and combined CL and CR groups, based upon *N*'s of 10, 10, and 20. The mean number of "same" response is graphed for each stimulus. For illustration of expected effects upon slope and symmetry of response curves, responses are plotted against stimulus magnitude measured on the physical dimension.

The most dramatic difference among the curves is one of slope, with both halves of the curve for the center stimulus range markedly steeper than those for the off-center ranges. An approximate index of slope for each *S* is taken as the sum of "same" responses to the two most extreme stimulus values within his stimulus range. Considering the median of all 40 scores, it is found that 17 of 20 *Ss* in the off-center range (LC and RC) and 3 of 20 *Ss* in the center range (CL and CR) exhibit scores above that median. An exact test of the significance (7) of differences between the indices for center vs. off-center ranges ($N_1 = N_2 = 20$) yields a two-tailed $P < .001$, and indicates the stability of the differences in slope between sets of curves for individual *Ss*. That this result occurs even with off-center stimulus units three times as large as center stimulus units attests to the importance of stimulus range as a determinant of the slope of response gradients.

The results of Fig. 1 also appear consistent with the prediction regarding skewness of response curves. The right half of the LC curve appears to drop with steeper slope than the left half, and the reverse is true of the RC curve. For individual *Ss*, 7 of the 10 in the LC group and 6 of the 10 in the RC group exhibit the predicted direction of skewness.

Central Tendency.—The independent variables considered with respect to Central Tendency are: (*a*) Order (whether a given stimulus range was presented as the initial or final condition); (*b*) Position (11 stimuli in each stimulus range employed); and (*c*) Cycle (one of five successively

presented blocks of 11 trials). In Fig. 2 are presented the mean curves generated in Test 2 by each of the four experimental groups. A comparison of these four curves with the three curves generated in Test 1, as presented in Fig. 1, appears to indicate that the mean response gradient for each of the groups is a function of its prior experimental experience with stimuli.

To simplify the analysis of differences between mean response gradients, there is tested the assumption that the LC and RC groups and the CL and CR groups should differ only as mirror images. After signs of the nominal stimulus positions are reflected (reversed) in both stimulus ranges for members of the RC group (henceforth labeled $\overline{RC}$), and in the off-center stimulus range for members of the CR group (henceforth labeled $\overline{CR}$), there are tested the significance of mean differences between LC and $\overline{RC}$ (in both ranges) and between CL and $\overline{CR}$ (off-center), for all dependent variables to be treated in subsequent analyses. For this test, use is made of the analysis of variance design for repeated measurements on the same *Ss*, labeled Type I by Lindquist (**8**, pp. 267–273). Because in no case did differences approach statistical significance, for subsequent analyses the four experimental groups are combined into two, reflecting order of conditions: the LC and $\overline{RC}$ group and the CL and $\overline{CR}$ group.

Because Position is defined in terms of different units in the center and off-center stimulus ranges, two Type I analyses of variance, one for each stimulus range, are carried out on the frequency of "same" response over five cycles at each stimulus position, to determine the effect of Central Tendency. The possible range of individual scores is 0 to 5. All 11 stimulus positions are included in the analysis for the off-center range. Because of low-frequency scores for the three positions at both extremes in the center range (− 5, − 4, − 3; and 3, 4, 5 mm.), assumptions underlying the analysis of variance are not met, and those data are excluded from the center range analysis, though the positions are represented in the figures.

Table 2 indicates, both for the central five stimulus positions of the center range, and for the 11 stimulus positions of the off-center range, that the Position by Order interaction effects which would reflect Central Tendency are significant. (The combined mean curves which were involved in the tests are not presented formally, but may be developed readily from the component curves appearing in the figures.) The off-center mean response curve for *Ss* having prior experience with the center range of stimuli (CL and $\overline{CR}$, Fig. 2) is displaced systematically toward the center range, as contrasted to the off-center mean response curve

TABLE 2

SUMMARIES OF ANALYSES OF VARIANCE FOR POSITION EFFECTS

Source	Center Range		Off-Center Range	
	df	Mean Square	*df*	Mean Square
Order	1	2.64	1	2.94
Error (b)	38	3.98	38	2.86
Total between-*Ss*	39	—	39	—
Position	4	45.59**	10	24.15**
Position × Order	4	8.06*	10	11.65**
Error (w)	152	2.83	380	2.80
Total within-*Ss*	160	—	400	—
Total	199	—	439	—

* Yields *F* significant beyond .05 level.
** Yields *F* significant beyond .001 level.

of *Ss* having no prior experience in the experiment (LC and $\overline{RC}$, Fig. 1). For the five stimulus positions in the center range, the mean response curve of the LC and $\overline{RC}$ group is displaced systematically in the direction of the off-center stimulus range involved in their prior experience, when contrasted with the mean response curve for the CL and CR group which had no prior experimental experience. The significant Position effects found in both ranges indicate that stable response gradients were obtained by individuals.

In the preceding analysis, the possible importance of Cycle has been ignored because the variable, frequency of "same" response, could take only the values zero or one for each test stimulus in each cycle, precluding the inclusion of Cycle and Position as independent variables in the same analysis. Cycle could effect Central Tendency as a result of either of two compensatory processes: its reduction due to increasing experience with the second stimulus range, or its enhancement due to negative time-order error (**2**, pp. 305–311). To study the effect of Cycle there is established a dependent variable which can take any one of a number of possible values: the arithmetic mean stimulus position value for each cycle, calculated only from those stimulus positions which called forth "same" responses. In two distinct Type I analyses of variance, one for each stimulus range, neither Cycle nor Order by Cycle effects approach significance. To the extent that the crude index is adequate, it is unlikely that systematic differences in performance over cycles account for the obtained results.

Summary

The aim of the present experiment was to investigate the effects of psychological vs. physical stimulus units and central tendency, determined by earlier experience with the stimulus dimension, upon generalization-discrimination gradients. A stimulus dimension of visual angle and a voluntary motor response were employed.

Four experimental groups, each of 10 adult human *Ss*, were employed. Each *S* was presented with two tasks: (*a*) training to respond to a zero angle stimulus, followed by testing for response to the training stimulus and to 10 stimuli symmetrically positioned around zero at intervals of 1′ 36″ (in the center stimulus range); and (*b*) training to respond to a nonzero angle stimulus, either a positive angle (in the right stimulus range) or a negative angle (in the left stimulus range) markedly different from zero, followed by testing for response to the training stimulus and to 10 stimuli symmetrically positioned around it at intervals of 4′ 48″. The four groups differed with respect to: the order of the tasks presented (whether a given stimulus range was presented first or second); and the position of the off-center training stimulus (either positive or negative).

Major results are as follows: (*a*) As expected from the assumed relationship between psychological and physical stimulus units, individual response gradients for *Ss* initially presented stimuli in the center range exhibited steeper slopes ($P < .001$) than those for *Ss* initially presented an off-center stimulus range. (*b*) As expected from the Central Tendency concept, response gradients of individual *Ss*, both in the center and off-center stimulus ranges, were displaced significantly in the direction of their previous experience with the stimulus dimension ($P < .05$, $P < .001$).

References

1. Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52–61.
2. Guilford, J. P. *Psychometric methods.* (2nd Ed.) New York: McGraw-Hill, 1954.
3. Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
4. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.
5. Hull, C. L. *A behavior system.* New Haven: Yale Univ. Press, 1952.
6. Johnson, D. M. The central tendency of judgment as a regression phenomenon. *Amer. Psychologist*, 1952, **7**, 281. (Abstract)
7. Latscha, R. Tests of significance in a 2×2 contingency table: extension of Finney's table. *Biometrika*, 1953, **40**, 74–75.

8. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
9. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
10. Philip, B. R. Generalization and central tendency in the discrimination of a series of stimuli. *Canad. J. Psychol.*, 1947, **1**, 196–204.
11. Raben, M. W. The white rat's discrimination of differences in intensity of illumination measured by a running response. *J. comp. physiol. Psychol.*, 1949, **42**, 254–272.
12. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

(Received June 20, 1955)

Journal of Experimental Psychology
Vol. 52, No. 1, 1956

HABIT STRENGTH AS A FUNCTION OF DRIVE IN A BRIGHTNESS DISCRIMINATION PROBLEM

EUGENE EISMAN

University of California at Riverside

ADELE ASIMOW AND IRVING MALTZMAN

University of California at Los Angeles

The present study deals with the relationship between learning and motivation. Cast in terms of Hull's system an attempt was made to determine whether ${}_sH_R = f(D)$. In a recent paper Spence (7) outlined the conditions necessary for a successful test of this relationship and indicated the shortcomings of certain previous studies. No extension of his analysis will be made here although the literature indicates a great deal of confusion and contradictory data.

The complete factorial design, a necessary condition for this test, as outlined by Spence, was utilized. Thus three groups of rats were trained under one of three levels of hunger deprivation, respectively. At the completion of training each main group was divided into three subgroups, which were then tested under one of the drive conditions of the three main groups. Such a design yields a 3×3 table, the rows representing drive during initial learning, the columns drive during the test. If the scores which are entered are test scores, and drive differences along the rows are equated through the factorization of the drive variable, any difference in the row means must represent some residual effect of the initial training. This difference, if present, is therefore a reflection of learning (${}_sH_R$).

Method

Apparatus.—The discrimination apparatus used in this experiment was adapted from Grice (3). It was of ¼-in. plywood, covered with ½-in. hardware cloth. The design and dimensions were as shown in Fig. 1. The starting boxes, choice area, and adaptation goal boxes were painted a flat, intermediate gray. The two outer goal boxes of the training unit were painted in white enamel, and the middle box in flat black. All of the guillotine doors were gray with the exception of those surfaces facing the inside of the black and the white boxes, respectively. These surfaces followed the color of the box.

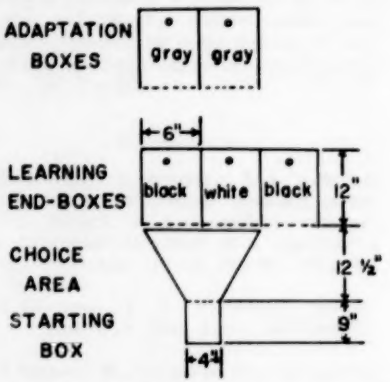

FIG. 1. Discrimination apparatus. The black-white-black boxes were replaced by the gray boxes during adaptation.

In addition to the general room illumination, a 60-w. lamp was centered 48 in. above the goal boxes.

Subjects.—The *S*s were hooded, male rats from the colony maintained at the University of California, Los Angeles, and were between 90 and 120 days at the start of the experiment. Sixty *S*s were started but 7, 6, and 4 were lost from the 4-, 22-, and 46-hr. drive groups, respectively. The results, threfore, are based on data obtained from 43 *S*s.

Maintenance.—Three levels of food deprivation (4-hr., 22-hr., and 45½-hr.) were utilized.

On the day *Ss* were taken from the stock house, they were divided into three equal groups, at random, and assigned to one of these drive levels. For the first 24 hr. all *Ss* were fed Friskies pellets (which were used as the reward in the subsequent phases of the procedure) ad libitum. On Day 2 the maintenance schedule was begun. The 22-hr. and 46-hr. groups were fed for 1 hr., 1 hr. after the time that their experimental sessions were subsequently begun. One-half hour was allowed for running time and a .5-hr. interval between the end of running and the beginning of feeding was imposed. A period of 22 hr. elapsed between the end of feeding and the beginning of the experimental session for the 22-hr. group. The 46-hr. group, of course, was fed only every other day and therefore 46 hr. elapsed between these two points. The 4-hr. group was fed for 1 hr. beginning 5 hr. before the scheduled running time. Four hours elapsed between the end of feeding and the start of running. To equalize intervals between experimental sessions (since the 46-hr. group could only be run once during a 48-hr. period) both the 4-hr. and the 22-hr. groups were also run only every other day.

On Day 22 of the procedure, four days before the learning phase began, the schedule was interrupted for 24 hr. of ad libitum feeding to allow some relief from the severe deprivation imposed on the 46-hr. group. To equalize conditions, however, the ad libitum feeding was extended to the other two groups as well. Figure 2 provides a graphic summary of the maintenance schedule.

Adaptation.—The adaptation phase covered a period of 25 days. On Day 1 the *Ss* were taken from stock, divided into three random groups as described above, and placed in their home cages, five *Ss* to a cage. Handling, marking, and institution of the maintenance schedule were accomplished on Days 2 through 8. Handling of a particular *S* took place only during that period which was set aside for the subsequent experimental treatment of that *S*, e.g., the 4-hr. group was handled 4 hr. after feeding.

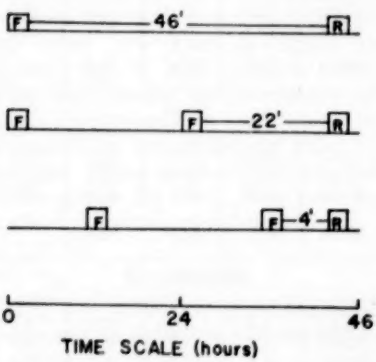

FIG. 2. Maintenance schedule. The 1-hr. feeding period is indicated by *F* and the 1-hr. experimental period by *R*.

One-half hour of group exploration of the discrimination apparatus was given on Days 9, 11, and 13. The gray end boxes were used. On Days 9 and 11 there was no food in the apparatus. On Day 13 there was a quantity of food in the cup in each box. Free trials were started on Day 15 and continued throughout the adaptation period. On Day 15, 2 trials were given, then 3, 6, 6, and 6 on Days 17, 19, 21, and 24, respectively. The *S* was placed in the starting box, the guillotine door raised, regardless of orientation, and the door to the end box lowered as soon as *S* was entirely in the box, i.e., a response was considered completed only when all four feet were within the box. Time was taken from the lifting of the starting door to the lowering of the end-box door. Food was in both cups during these trials. During the first seven trials the large quantity of food in the cup was gradually reduced to one pellet, which remained constant from then on.

Original learning.—The black-white discrimination learning was begun on Day 26. The procedure was the same as during adaptation except that the black-white boxes were used and a pellet was placed in only the white cup. Six trials per day were given for all groups. A decided black preference was observed in pilot work and therefore all *Ss* were trained white-positive during this phase. The criterion of learning was 14 correct out of 16 trials. The right-left placement of the reward and therefore the placement of the white box was randomly determined. Trials were rotated with the intertrial interval approximately 5 min. The *Ss* were detained in the end box for 20 sec. on an incorrect run. A noncorrection procedure was adopted. The *Ss* were kept in individual retaining cages between trials. These were different from both the home cage and the end box in construction, color (brightness), etc.

Drive shift.—After the criterion was reached on the original problem, drives were shifted and *Ss* were trained to the reversal, i.e., black-positive. Between the white-positive and black-positive training *Ss* were fed ad libitum for 48 hr. and then adapted to the new drive for 48 hr.—one-third of *Ss* retained on old drive, one-third shifted, randomly, to each of the other drives and reversal training to black-positive was to the same criterion as in original learning.

TABLE 1

PERFORMANCE MEASURES IN ORIGINAL LEARNING

Drive	*N*	Errors	Reinforcements	Trials
4-hr.	14	19.9	25.6	45.5
22-hr.	14	17.2	27.1	44.4
46-hr.	15	11.9	20.5	32.3

Tests of Significance

Errors:

$F = 6.69; P < .01$

$t_{4-22} = 1.18; P > .05$

$t_{22-46} = 2.40; P < .03$

Reinforcements:

$F = 7.79; P < .01$

$t_{4-22} = .83; P > .05$

$t_{4-46} = 2.91; P < .01$

Trials:

$F = 8.26; P < .01$

$t_{4-22} = .31; P > .05$

$t_{22-46} = 3.33; P < .01$

Four measures were obtained during reversal training: errors, reinforcements, trials to the criterion, and resistance to extinction (*n*). This measure was defined as the number of trials on which *S* returned to the formerly correct side (white) before the first run to the other side (black).

RESULTS

Data for the initial learning are presented in Table 1. A summary of the various tests of significance is also included. With respect to each of the three measures, a simple analysis of variance indicated a significant difference among groups. The probability values are all beyond the .01 level. The *t* tests shown were based on the use of the common error estimate of the variance analysis. It may be seen that the 4- and 22-hr. groups did not differ significantly with respect to any one of the three measures. The 46-hr. group did differ, however, from each of the other groups at the .03 level or better. Thus, the 46-hr. group required fewer errors, trials, and reinforcements to reach the criterion than either of the other two groups.

Data on the reversal learning are presented in Table 2. This table is composed of five subtables, each arranged in the form of a 3 × 3 table with the columns representing drive during reversal and the rows drive during original learning. All cell as well as border entries are mean values

TABLE 2

PERFORMANCE MEASURES IN REVERSAL LEARNING

Original Drive	Reversal Drive			Total
	4-hr.	22-hr.	46-hr.	
Number of Cases				
4-hr.	5	4	5	14
22-hr.	5	5	4	14
46-hr.	5	6	4	15
Total	15	15	13	
Errors				
4-hr.	17.6	14.2	14.6	15.6
22-hr.	16.4	23.4	14.5	18.4
46-hr.	20.4	21.0	20.5	20.7
Mean	18.1	20.0	16.4	
Reinforcements				
4-hr.	23.0	21.0	20.2	21.4
22-hr.	22.2	25.0	21.0	22.9
46-hr.	20.4	20.2	20.0	20.2
Mean	21.9	22.0	20.4	
Trials				
4-hr.	40.6	35.2	34.8	37.0
22-hr.	38.6	48.4	35.5	41.2
46-hr.	40.8	41.2	40.5	40.9
Mean	40.0	42.0	36.8	
Extinction (*n*)				
4-hr.	2.6	4.5	3.9	3.6
22-hr.	4.4	5.2	4.5	4.7
46-hr.	8.8	6.2	8.2	7.6
Mean	5.3	5.4	5.4	

except for the subtable indicating the number of cases. In the latter, the borders indicate the total number of *Ss* in each of the three main groups during reversal and original learning, respectively, and the cells the number of *Ss* in each of the nine subgroups composing these main groups.

Four separate analyses of variance[1] were performed for each of the response measures. None of the *F*'s associated with the column main effects were significant, nor were any of the column-by-row interactions. Thus, drive during reversal had no differential effect on learning. The row main-effects analyses did, on the other hand, reveal a significant difference for one of the response measures. As in the case of the column data, neither errors ($F = 2.01$, $P > .05$), reinforcements ($F = 1.23$, $P > .05$), nor trials ($F = .70$, $P > .05$) varied during reversal as a function of the drive during original training. Resistance to extinction, however, did reflect differences as a function of original drive ($F = 16.5$, $df = 2$ and 34, $P < .01$). Resistance to extinction between groups trained under 4- and 22-hr. of drive was not significant ($t = 1.49$); whereas, the differences between the 46-hr. group and each of the other two were significant at beyond the .01 level ($t_{22-46} = 3.98$). This was similar to the results of original learning. In both cases the 4- and 22-hr. groups did not differ on any one of the measures whereas the 46-hr. group did differ from both on each of the obtained indicators.

Discussion

The first point that bears discussion is the lack of difference between the 4- and 22-hr. groups during both original and reversal learning. A possible explanation involves the maintenance procedure. Both groups were fed 1 hr. during a 24-hr. period. The 4-hr. group was run 4 hr. after feeding, the 22-hr. group was run 22 hr. after feeding. Both were deprived for a period of 23 hr. out of 24. If drive were a function of two parameters, the total deprivation time and hours of deprivation at the time of running ($\hat{h}$) and $\hat{h}$ reached an asymptote at or before 4 hr. then both the 4-hr. and the 22-hr. groups would have had the same degree of drive and no differences would have been expected. That some sort of two-parameter definition of drive is necessary is indicated by work now in progress (**2**).

The interpretation of the effects of drive differences during reversal on reversal learning, as presented in Table 2, is relatively simple. Lack of significance on any one of the response measures indicates that drive during reversal had no differential effect on learning. Why it should be ineffective during reversal learning when it was effective during original learning is a more difficult question. It may have resulted from the fact that at the beginning of reversal training, all *Ss* had fairly strong incorrect habits as a result of the original training. The multiplication of these incorrect habits by high *D* could cause more interference than when multiplied by low *D*. The 46-hr. group on this basis would therefore be expected to do more poorly. It was seen, however, that high *D* had a facilitating effect on performance during original learning, and from this point of view the 46-hr. groups should have done better than the lower drive groups on the reversal learning. These two factors, one facilitating, the other retarding, may have canceled each other and thus produced apparent equality of learning with respect to drive differences during reversal.

The interpretation of the effects of drive differences during original learning on reversal learning presents a more complicated picture. Neither errors, nor

[1] The 3 × 3 analysis involved disproportional subgroup frequencies. Correction for disproportionality was accomplished by the method of Walker and Lev (**10**).

reinforcements, nor trials during reversal yielded significant differences as a function of drive during original learning, indicating the independence of habit strength and drive. Differences between means on resistance to extinction, however, were significant at beyond the .01 level, and this was primarily due to the greater resistance to extinction of *S*s trained under the 46-hr. drive. Since the groups were equated with respect to drive during reversal, only two factors could account for this difference. If *S*s trained originally under 46-hr. deprivation had acquired more ${}_{S}H_{R}$ for white than either of the other groups, this would be expected. According to Hull's most recent formulation (**5**), ${}_{S}H_{R}$ is assumed to be a function of the number of reinforcements. Referring to Table 1, it is seen that the 46-hr. group actually received significantly fewer reinforcements than the other groups. This should therefore have led to the development of less ${}_{S}H_{R}$ and consequently a weaker n. That this was not the case is here explained by the possibility that ${}_{S}H_{R}$ and D are not independent. If the increments in ${}_{S}H_{R}$ were greater under the higher levels of D than under lower, then the 46-hr. group could have acquired more ${}_{S}H_{R}$ for white during original learning, making their reversal learning more difficult, i.e., producing greater n for the 46-hr. group. The second factor that might have accounted for the "slower" learning of *S*s trained under 46-hr. deprivation, originally, is conditioned inhibition (${}_{S}I_{R}$). If the 46-hr. *S*s developed more ${}_{S}I_{R}$ for the black during original training, then they would have tended to avoid the shift to black after reversal. Referring to Table 1, it may be seen that the 46-hr. group had fewer errors, and also fewer total trials to criterion than either of the other two groups, implying that, if anything, their ${}_{S}I_{R}$ for black was less than for either the 4- or 22-hr. groups. This factor, then, should have worked against the direction of the obtained results. It was also the case that trials were spaced during the original as well as reversal phases, minimizing the accumulation of ${}_{S}I_{R}$. It is concluded therefore that ${}_{S}H_{R}$ is functionally related to the drive under which it is developed.

The question of why errors, reinforcements, and trials did not differ significantly, whereas n did differ can be answered only speculatively, on the basis of the results of this study. It may simply be that these measures are not sensitive enough to reflect small differences. Another more attractive possibility is that once the initial extinction occurs, there are no longer any differences to be measured. The groups, it should be remembered, were equated for all variables except drive during original training. If the effects of drive during original learning are equalized during the initial period of reversal learning, i.e., during the 3.6, 4.7, and 7.6 successive errors at the beginning of reversal, then no differences in learning after that period should be expected. This, of course, was the case. The extreme similarity in relearning can more easily be observed if the extinction errors are subtracted from the total errors or from the trials to relearn. The scores for errors after this subtraction for the 4-, 22- and 46-hr. groups, respectively, are 12.0, 13.7, and 13.1. For trials they are 33.4, 36.5, and 33.3. Considering the small N and the usual variability of rat data, the differences between these figures are indeed small.

This interpretation is further supported by the results of a recent study (**1**). Using a straight alley, high- and low-drive groups of rats were given 14 training trials to food. After training, each group was divided into two subgroups and extinguished under one of the two drive levels of the training period, respectively. Differences in trials to extinction, as a function of drive during training, were not significant. Differences in running speed attributable to drive during training, for the first six extinction trials, however, were significant. The latter data would indicate, therefore, the dependence of habit strength on the drive during learning, whereas the former would indicate the independence of the two

constructs. Two interpretations were offered as a means of explaining this ambiguity. One questioned the validity of Hull's use of both amplitude (running speed) and resistance to extinction as alternative measures of response strength. The second suggested that the high-drive *S*s learned responses which were different from those learned by *S*s trained under low drive. It is felt that the interpretation given in the present study is more parsimonious than the two offered by Campbell and Kraeling. Referring again to that study, it was reported that the running-speed differences during extinction were most disparate during the first few extinction trials, and it was shown that by Trial 10 there were practically no differences between *S*s trained under high and low drive, respectively. In other words, differences between groups were only apparent during the initial stages of extinction. If a more severe criterion of extinction were employed and/or a greater number of trials than the first six extinction trials were used, it is probable that even the amplitude measure (running speed) would have yielded insignificance. The point is that the amplitude measure yielded significance not because it measured a different aspect of response strength from trials to extinction or that different responses were learned under high as compared with low drive, but that the amplitude measure reflected differences due to training drive level which were only present during the early stages of the extinction period. In certain other studies of similar design (**4, 6, 8, 9**), trials to extinction were also used and yielded negative results. That is to say, no differences were obtained, and it was concluded that habit was independent of drive. The implication of the Campbell and Kraeling interpretation would be that the lack of agreement between their study and the others is attributable to the use of an amplitude measure in theirs and a resistance-to-extinction measure in the others. It is suggested here, however, that the lack of agreement is due to the use of a measure which in the one case reflected differences during the early phases of extinction and in the other cases a measure which was insensitive to these intial differences. It is not that amplitude and resistance to extinction measure reaction potential differently or that they measure different aspects of it, but that they measure or reflect reaction potential at different points in the learning process.

Summary

Three groups of hooded, male rats were trained to run to white in a black-white discrimination apparatus under 4, 22, and 46 hr. of food deprivation, respectively. After reaching the criterion of learning, on white-positive, drives were shifted and the *S*s retrained to black-positive. The drive shift involved the division of each main group into three subgroups, each corresponding to one of the drive conditions of the three main groups, respectively.

No significant differences were obtained between the 4- and 22-hr. groups either during the original, white-positive or during the reversal, black-positive learning phases. An hypothesis involving a two parameter definition of drive was offered to explain this lack of difference.

The 46-hr. group was significantly superior to the lower drive groups during original learning with respect to errors, reinforcements, and trials to the criterion of learning. The 46-hr. group showed greater resistance to extinction than either of the low-drive groups during the reversal phase, although it did not differ with respect to errors, trials, or reinforcements to the criterion of relearning. This was explained on the basis of the initial extinction of differences at the beginning of reversal learning and it was concluded that habit strength bears some functional relationship to drive during learning.

References

1. Campbell, B. A., & Kraeling, D. Response strength as a function of drive level during training and extinction. *J. comp. physiol. Psychol.*, 1954, **47**, 101–103.
2. Eisman, E. An investigation of the parameters defining drive. *J. exp. Psychol.*, in press.
3. Grice, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, **38**, 1–16.
4. Hillman, B., Hunter, W. S., & Kimble, G. A. The effect of drive level on the

maze performance of the white rat. *J. comp. physiol. Psychol.*, 1953, **46**, 87–89.
5. HULL, C. L. *Essentials of behavior.* New Haven: Yale Univ. Press, 1951.
6. KENDLER, H. H. Drive interaction: II. Experimental analysis of the role of drive in learning theory. *J. exp. Psychol.*, 1945, **35**, 188–198
7. SPENCE, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.
8. STRASSBURGER, R. C. Resistance to extinction of a conditioned operant as related to drive level at reinforcement. *J. exp. Psychol.*, 1950, **40**, 473–487.
9. TEEL, K. S. Habit strength as a function of motivation during learning. *J. comp. physiol. Psychol.*, 1952, **45**, 188–191.
10. WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

(Received June 30, 1955)

QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Founded in 1948

Published Quarterly in February, May, August, and November

for the

EXPERIMENTAL PSYCHOLOGY GROUP

Editor: D. Russell Davis, University of Cambridge

Subscriptions from the United States
should be entered now for the next year.
30 shillings per year (current rate $4.20).

Order from

W. Heffer & Sons, Ltd.
Cambridge, England